



**KTH Information and  
Communication Technology**

# **The Opinion Evaluation Network: Ranking Imprecise Social Interactions**

A Master thesis submitted in partial fulfillment of the requirements for the  
Degree of Master of Science (Teknologi MagisterExamen) in Interactive Systems Engineering

**NIKOLAOS THEODOROS KORFIATIS**

Department of Computer and Systems Sciences (DSV)  
Royal Institute of Technology (KTH)  
Stockholm, Sweden

Stockholm, December 2005

### **Abstract**

Current models of ranking in Information Retrieval (IR) are somewhat “blind” to the consideration of the social context that surrounds an information resource as a parameter that affects the precision of the ranking in the query results. On the other hand that social context is depicted upon the relational ties of the affiliated social entities (authors) thus is something that cannot be measured and quantified accurately by back-link and citation based models. In this thesis we adopt an imprecise modeling approach of the depicted relational ties using the paradigm of fuzzy sets as to express partial degrees of membership depicted on the concept of “opinion” as an input to a model that considers both the informational (hyperlink) and social (relational) context of the information resources as to provide better ranking of the retrieved results with respect to both contexts. A formalization of the algorithm and a validation of the model using simulation is given as a proof of concept along with discussion of the obtained results.

# Contents

<b>Contents</b>	<b>i</b>
<b>List of Figures</b>	<b>iii</b>
<b>List of Tables</b>	<b>iv</b>
<b>1 Introduction</b>	<b>1</b>
1.1 Background and Motivation . . . . .	1
1.2 Aim and Research Question . . . . .	2
1.3 Targeted Audience . . . . .	3
1.4 Limitations . . . . .	3
1.5 Thesis structure . . . . .	4
1.6 Acknowledgments and Method of Work . . . . .	5
<b>2 Opinions, Ranking and Social Networks</b>	<b>7</b>
2.1 Introduction . . . . .	7
2.2 Analytical Models of Social Studies . . . . .	8
2.3 Sociometric Theories . . . . .	10
2.3.1 Balance Theory . . . . .	10
2.3.2 Cliques and Cohesion . . . . .	11
2.4 Social Network Analysis . . . . .	12
2.4.1 Formal Representations . . . . .	13
2.4.2 Design of Sociometric Studies . . . . .	14
2.4.3 Applications of Social Network Analysis . . . . .	16
2.5 Indices of Social Ranking . . . . .	17
2.5.1 Prestige . . . . .	18
2.5.2 Centrality . . . . .	19

<b>3</b>	<b>Ranking in Information Retrieval</b>	<b>21</b>
3.1	Background . . . . .	21
3.2	The Web as a graph . . . . .	22
3.3	Citation and Back-link Models . . . . .	23
3.3.1	Hubs and Authorities . . . . .	23
3.3.2	The PageRank Citation Ranking . . . . .	25
3.4	Semantic Web . . . . .	26
3.4.1	The FOAF Vocabulary . . . . .	28
<b>4</b>	<b>A fuzzy approach on the representation of social ties</b>	<b>31</b>
4.1	Traditional Representations . . . . .	31
4.1.1	Strength . . . . .	33
4.1.2	Evidence . . . . .	33
4.2	Fuzzy representations of Strength and Evidence . . . . .	33
4.2.1	Representation of Sets . . . . .	34
4.2.2	Representation of Membership . . . . .	35
4.3	Aggregation . . . . .	36
4.3.1	The Ordered Weighted Average Operator . . . . .	36
4.4	Opinion as an aggregation function . . . . .	36
<b>5</b>	<b>OpinionRank: A fuzzy adapted ranking model</b>	<b>39</b>
5.1	Qualitative Exploration . . . . .	39
5.1.1	OpinionRank . . . . .	40
5.2	Implementation . . . . .	40
5.2.1	Algorithm Implementation . . . . .	40
5.3	Representing Fuziness . . . . .	41
5.3.1	Set and Variable Definition . . . . .	41
5.4	Normalization . . . . .	44
<b>6</b>	<b>Validating OpinionRank</b>	<b>45</b>
6.1	Why Simulation . . . . .	45
6.1.1	Shortcomings . . . . .	45
6.2	Model Construction . . . . .	46
6.2.1	Input Parameters . . . . .	46
6.2.2	Data Collection and Visualization . . . . .	47

6.2.3	Running the model . . . . .	49
6.3	Discussion . . . . .	52
<b>7</b>	<b>Conclusions and Future Research</b>	<b>55</b>
7.1	General Conclusions . . . . .	55
7.2	Lessons Learned . . . . .	56
7.3	Future Research . . . . .	57
<b>A</b>	<b>Running the Simulation Environment</b>	<b>59</b>
<b>B</b>	<b>Index of Abbreviations</b>	<b>61</b>
<b>C</b>	<b>Literature Collected</b>	<b>63</b>
	<b>Bibliography</b>	<b>65</b>

## List of Figures

1.1	Thesis Structure . . . . .	4
2.1	Social Science Research Models . . . . .	8
2.2	Cartwright-Harary triples. Notice the directional connections in the second triple (signed triple) . . . . .	11
2.3	A bidirectional sociogram along with its adjacency matrix . . . . .	13
2.4	The Three basic types of rankings in a network: Star, Circle and Line . . . . .	17
3.1	Typical architecture of a web search engine as a user-query interface . . . . .	22
3.2	Hubs and Authorities contained in a topic . . . . .	24
3.3	Network of pages in the current and the semantic web . . . . .	27
4.1	Basic Operations on the fuzzy sets A and B and the value of their membership function along with the transition from the one fuzzy set to the other . . . . .	35

6.1	Entering simulation parameters . . . . .	46
6.2	Simulation Model . . . . .	47
6.3	The sociomatrix of the resulted network after 10 rounds. The black boxes indicate whether this connection exists or not. Visualization was done using the Pajek software for network analysis ([Batagelj and Mrvar, 2002]) . . . . .	48
6.4	Social network view of <i>OpinionRank</i> . . . . .	52
6.5	Opinion Evaluation depicted by the hyperlink context on the network of the web pages	53
6.6	Opinion Evaluation depicted by both the hypertextual and social context of the web pages and their affiliated authors . . . . .	54
6.7	Evaluation Plot. The blue series represent the values obtained from the original PageRank ranking . . . . .	54
A.1	Launching the simulation environment . . . . .	59

## List of Tables

2.1	Contextualized SNA studies in the Social Sciences . . . . .	17
3.1	Basic Elements of the FOAF Vocabulary and their type relational tie . . . . .	30
4.1	Traditional and Fuzzy(Linguistic) Expressions of Directional Relations in a Social Network . . . . .	32
6.1	Ranking of the node set obtained from PageRank . . . . .	50
6.2	Ranking of the node set obtained from OpinionRank, notice the change of the ranking	51

# Chapter 1

## Introduction

### 1.1 Background and Motivation

Society is a structure that is formed from a set of interpersonal and institutional relationships that affect the way people communicate, collaborate and coordinate their activities. Moreover the strength of those relationships and the patterns that are formed between people and groups affect the behavior of individuals and the formation of groups inside the society. On the other hand relationship means communication between social entities that produces and consumes information which depending on the context and the situation is of great importance. Since the advances of Internet and web technology have transformed the communication channel from a temporal to preserving one ([Licklider et al., 1968]), the research question comes on how to evaluate this information to the needs and preferences of the person requesting it.

Furthermore since information is produced and consumed by humans apart from the content of the information resource one also needs to examine the social context ([Brown and Duguid, 2002]) of this information and the trustworthiness of its creator in order to derive a safe decision. Current models on information retrieval (IR) and web search engine retrieval ([Faloutsos, 1985]) are somewhat blind to the above consideration since they try to measure credibility based on the citation and back-link characteristics of the resource without taking into account the strength and the affective value of this citation. Let us consider for instance the evaluation of a scholar based on the number of citations that other scholars give to his work. Current models in scientometric research (for a review see [Leydesdorff, 2001]) take into account a bivalent association that denotes whether the citation exists or not, without examining the strength (e.g. is it noted as a classical example ?) or the affection that it characterizes the reference (eg. is it noted as an example of good or bad work ?)

In society evaluation is depicted on the “opinion” of an individual about someone else. Meriam-Webster defines an opinion as “*a view, judgment, or appraisal formed in the mind about a particular matter*” which as a definition provides as the first pillar onto which we build our argumentation. However “opinion” is not always depicting an accurate judgment since evaluation is affected by several properties including affective relations. Therefore a need for combination of multiple opinions needs to be done in order for the more accurate evaluations to emerge. That is an opinion evaluation network where evaluation emerges rather than being attributed.

In this thesis we try to evaluate, using an imprecise modelling approach, the trustworthiness and the credibility of the creator/author of the information resource using input from his/her social context as an extra parameter to information retrieval models that facilitate the extraction of the most relevant and respectful/popular information resources. We denote the concept of “opinion” as a non reciprocal characteristic of a social entity that reflects its credibility from a set of relationships formed inside a network. The imprecise evaluation of this opinion leads us to the extraction of the relevant trustworthiness that is attached to the information resource authored by the social entity.

## 1.2 Aim and Research Question

The aim of this thesis is to elaborate on the design of a measure of credibility based on imprecise expressions of social criteria that can be inferred from the social context of the author to whom that information resource is affiliated with, as well as to express attributed trustworthiness denoted from the social context to that information resource.

Furthermore the research question addresses the forms of representation of social criteria that can be achieved using advances of web technology such as the semantic web. To this end an algorithm will be developed as a proof of concept and a validation using a simulation model will be carried in order to investigate:

- (a) How can we model and represent relational ties by using imprecise representations of their characteristics ?
- (b) How biased can the results of the standard web metrics be by consideration of social criteria expressed in imprecise form and
- (c) What kind of connection exists between the two contexts (is there any reciprocity between them ?).

## 1.3 Targeted Audience

Our targeted audience spans the intersections of the branches of mathematical and analytical sociology along with the field of information retrieval and in particular search engine research. Fuzzy set theory is used as an input to information retrieval models that support the evaluation of social theories such as social influence and support theory ([Friedkin, 1998]). Both researchers on social and information sciences are the main target of this thesis since we try to model the aforementioned problem and research question using inputs and models of computer science and sociology. On the other hand, interaction designers and designers of information services in general will consider this thesis useful regarding the inputs expressed and communicated in the representation of ties and group structure as an input to requirements analysis and in particular the analysis of organizational dependencies necessary in cases such as Computer Supported Cooperative Work (CSCW). Furthermore the flexibility of the world wide web as a platform to support spontaneous collaboration in communities of practice using “social”<sup>1</sup> software such as folksonomies ([Mathes, 2004]) and wikis ([Leuf and Cunningham, 2001]) has implied the demand for studies that can expand the merit and the effectiveness of such type of network based interaction in order to lead to more effective design.

## 1.4 Limitations

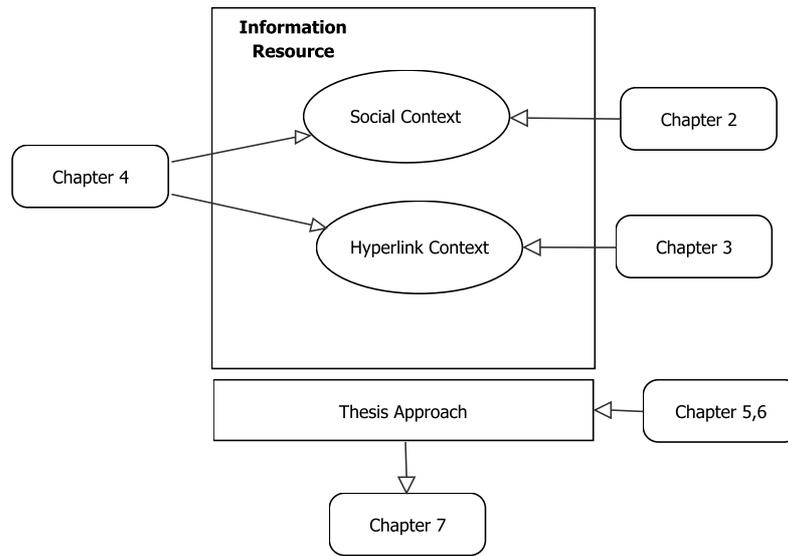
The basic axiom that has motivated us for the writing of this thesis is that social ties cannot be expressed accurately with models based on formal logic and in particular crisp set theory. Therefore representation of such models with richer expressions is needed. However, such properties which belong to non-distinct boundary sets of values, may not be characteristic of a social tie, such as e.g., in the cases of:

**Transaction** In case of Economic Networks a transaction is an event that has only two instances (Happened or not) therefore an imprecise expression is out of score with however considering the fact that there might be some exceptions

**Membership** A person is affiliated in a group (ethnic, social, etc) where properties of the group are depicted on that person. For instance if we try to classify the persons based on racial categories we can have membership only to one of them etc.

---

<sup>1</sup>The terms social software refers to web based software that supports spontaneous collaboration over the WWW



**Figure 1.1:** Thesis Structure

Furthermore depending on the context of the study of affiliated properties may not always demand the explicit modeling.

## 1.5 Thesis structure

Figure (1.1) provides an overview of the structure and the relations between the thesis chapters. More specifically the thesis is structured as follows:

**Chapter 2** provides a background on social and analytical models that extract metrics of social *status* and in particular provides an argumentation on the selection of the network paradigm as a research method for our thesis.

**Chapter 3** provides an overview of information retrieval models such as HITS and PageRank which are used to evaluate trustworthiness based on the hyperlink context of the resource.

**Chapter 4** gives an insight to the imprecise properties of relational ties and their representations using fuzzy models.

**Chapter 5** builds the evaluation algorithm and discusses the approaches taken to include the aforementioned parameters on the evaluation model.

**Chapter 6** conducts a simulation study and experiment using the proposed algorithm as a comparison metric.

**Chapter 7** draws conclusions and discusses future research that can extend the proposed model.

## **1.6 Acknowledgments and Method of Work**

The original motivation for this thesis was resulted after long hours of discussion with my supervisor Dr. Ambjörn Naeve head of the Knowledge Management Research Group(KMR) at the School of Computer Science and Communication (NADA) in the Royal Institute of Technology(KTH). Further development and ideas were communicated with KMR members Ass. Prof. Miguel-Ángel Sicilia and Elena Garcia from the University of Alcalá, Spain, as well as with Dr. Miltiades Lytras from the Research and Academic Computer Technology Institute (CTI) in the University of Patras, Greece. Earlier versions of this thesis were presented as a poster in the PROLEARN summer school that took place in Constantinople from 7 to 14,September 2005. The author wishes to acknowledge the input and discussions from fellow colleagues Michele Passin (Open University, London) and Riina Vuorikari (KUL - Leuven, Belgium) for their inputs and comments.

Endless hours of MSN and Skype Chatting have eliminated the geographical and social distance between us and have given great support to my inquiries and problems through the experimental and writing part of this work. The author wishes also to acknowledge the support provided by the examiner Dr. Harko Verhagen from the Department of Computers and Systems Sciences (DSV) for arranging the administrative issues and making this thesis possible.



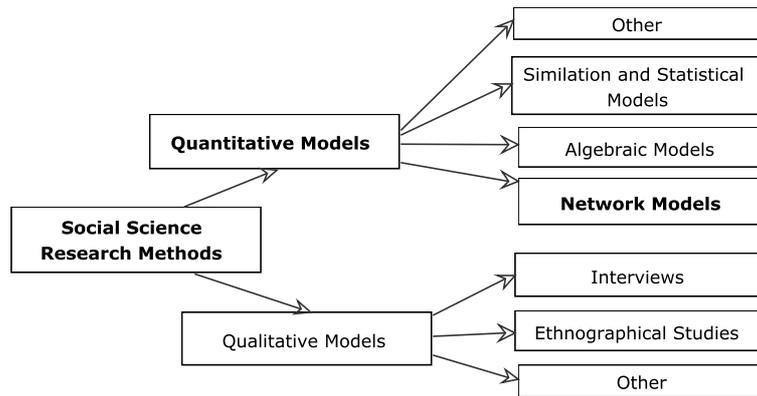
## Chapter 2

# Opinions, Ranking and Social Networks

### 2.1 Introduction

Sociology and the social sciences in general, targets to develop models and theories that will help researchers understand and explain behaviors as a consequence of membership and exposure to different situations and contexts that are formed through interactions between individuals and groups ([March and Lave, 1975]). These types of interactions often referenced in the literature as “Social Interactions” ([Kelley, 1971]) consist of a large typology of connections between social entities referenced to this thesis from now on as “relational ties”. In our approach “Social Interaction” refers to the communication interaction between two or more social entities thus becomes the prerequisite for the formation of an opinion of one entity by the other. Other interpretations of social interaction such us those of social actions is not a subject of this thesis.

Contextualized interpretations of relational ties become on great importance in the related fields of economics and political science where the way people interact is of great interest to those that try to construct and evaluate the trajectories that influence the individual and institutional entities. In political science for instance, people are called to evaluate the importance of and trust a person and its ability to represent their “crisis” on important matters affected by the decisions of that particular social entity. In Economics that type of evaluation is important when it comes to decide for the provision of goods and services by the entity that the others tend to evaluate positively. For example consider a modern scenario where one might want to find an individual or a company to provide a service. Apart from the formal dependencies that have to be fulfilled (e.g the ability to provide the service) there is also the question for *what types of quality will the provided service have*. Such social processes constitute a type of evaluation where an entity of the social structure is evaluated by the opinions of the others thus is being attributed



**Figure 2.1:** Social Science Research Models

a rank or status([Katz, 1953]). According to Raub ([Raub and Weesie, 1990]) in social contexts two types of ranking exist: **Explicit** (Formal) and **Implicit**(Informal). In explicit ranking it is declared who is the one who has the authority to command due to insignia, symbols or status that he/she is being attributed from the beginning. On the other side in implicit ranking is the *opinion and the behaviors of the surrounding entities* that facilitates who has the authority to command and decide.

Both implicit and explicit forms of ranking may exist depending on the purpose of the social group and the kind of interactions between the members. However our study is focused only on implicit forms of ranking where *status* is a social attribute that emerges rather than being set. In sociology there are several kinds of studies that tend to explain how status emerges, therefore several research models exist. Those models are analyzed further in the sections bellow.

## 2.2 Analytical Models of Social Studies

Analytical studies in sociology tend to explain not only the behavior but also the surrounding context that influences the individual to expose different forms of behavioral patterns through its interaction with the other members of the group [March and Lave, 1975, Moreno and Jennings, 1945]. To this end both qualitative and quantitative approaches have been applied to provide interpretations such as to explain performance of individual and groups under different contexts whereas experiments were carried out in order to examine the position of individuals under different contexts. As can be seen in the figure (2.1), both quantitative and qualitative research methods for the analysis of contextual data exist in sociology. In our approach the kind of data we wish to examine is *non-descriptive*, for instance the interaction in a social group cannot be

measured as a whole by using qualitative methods therefore in our study we consider the formal (quantitative) approach which consists of the following models:

**Algebraic Models:** Algebraic Models (for an overview see [Pattison, 1993]) are used to describe multiple properties of a social structure to the extent of combining and inferring relationships. Structural models based on algebraic operations are used to study influence and transitivity as well as set of properties of the examined entities particularly known as social roles [Ross et al., 1977]

**Simulation and Statistical Models:** Statistical Models such as Multidimensional Scaling [Kruskal and Wish, 1978] try to interpret interactions using the clusterability of the set containing the social entities as an indicator of prominence. Simulation models try to reproduce the formation of structures in order to provide data useful for the analysis of the properties of entities and groups, by the statistical and algebraic model.

**Network Models:** Network Models are formalized using Graph Theory [Harary, 1969] where principles such as the centrality index or the geodesic distances are used to interpret the directional or bidirectional interactions formalized through the model.

All three models are used extensively in the literature however in our research question the issue comes on which paradigm shall we follow to express trustworthiness as an outcome of interactions. Several sociologists have argued ([Raub and Weesie, 1990],[Kelley, 1971]) that since social interactions are subject to context, network topologies may be context-aware on the expressiveness of dependent and interdependent properties of social entities when interacting. Furthermore in our case we wish to examine the relevance of an information resource using a trustworthiness measure regarding how well ranked the author of this resource is on a social context. However network modeling compared to relational algebras and other algebraic models of analytical studies may lack the expressiveness required to capture the properties of those interactions. The mix of the above three models results a special kind of studies to which their objective is to measure the properties and the attributes of the social structure by using statistical, network and algebraic models. This studies are known in sociology as “Sociometric studies” and the relevant branch as “Sociometry”. Sociometry forms the core of “Social Network Analysis” which is the main research method that was used in this study. Furthermore apart from the formal methods, a sociometric study also take an input from theoretical motivations in order one to be able to provide interpretations for the results as well as inputs for further

analysis. The basic theories that are used extensively in the sociometric literature are analyzed in the section above.

## 2.3 Sociometric Theories

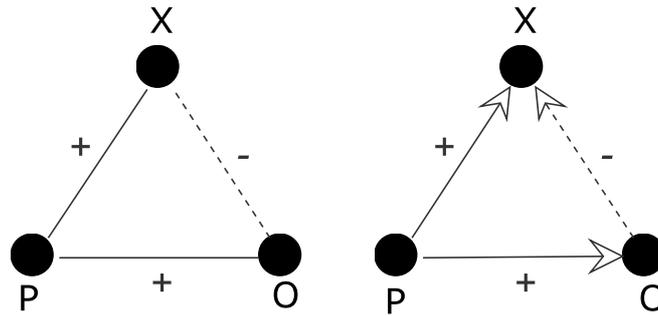
Sociometric theories form the basis for the constructions of models that measure and explain properties of entities and groups in interaction situations. The basic assumption behind sociometric theories is that social actors are interdependent and that the links among them have important consequences for every individual [Freeman, 2000]. Most celebrated among those is Balance theory and its extension of Structural Balance as well as the theory of clique formations and cohesion which tends to explain theoretically the way interactions are formed and clustered in a social group.

### 2.3.1 Balance Theory

Several sociometric interpretations or ranking originate from the field of social psychology where the impact of group processes is modeled upon the perceptions and the behavior of an individual [Aronson et al., 2004]. In those contexts the focus comes to unravel the individual that possesses the most central position on the group thus influence the others upon his/her choices. A series of theories has been developed with the most notable one the *Balance Theory* proposed by Heider [Heider, 1946]. Balance theory builds on the principle that when an individual falls into tensions with another member of the group then it tries to eliminate this tension through self-persuasion or persuasion of other entities. That means that the entity will try to break apart its interaction with the entity that falls into tension thus that entity will receive less choices/opportunities to interact.

Cartwright and Harary [Cartwright and Harary, 1956, Cartwright and Harary, 1970] formalized and extended Heider's theory by defining a special kind of network where they tried to represent structures of affective ties (interactions that contain an expression of emotions) known in sociometry as Person-Other-Object(X) or Cartwright-Harary triple (see figure 2.2). Affective ties can represent positive or negative emotions such as: praise or blame, love or hate and so forth. Furthermore Heider's original proposition of that affective or cognitive balance has been generalized to describe the group or "structural balance" which considers the overall affective ties that are formed in a group or social structure.

Balance theory and Structural Balance contribute to the decomposition of the social structure in substructures (clusters) where representations of dominance occur. For instance in an



**Figure 2.2:** Cartwright-Harary triples. Notice the directional connections in the second triple (signed triple)

“unbalanced” structure the one who dominates is the one who is the most prominent and influences the actions of the others via each relations. Unbalanced situations break cohesion and lead to the formation of cliques explained by the sociometric theories dealing with cliques and cohesion analysis.

### 2.3.2 Cliques and Cohesion

The graph theoretic definition of a clique has its qualitative interpretation in sociometric research as a discrete social structure (subgroup) shaped and contained within the structure of a social group ([Bron and Kerbosch, 1973]). The number of cliques contained in a group and their diameter is subject to the cohesion that characterizes the social structure. Cohesion depicts how strong the ties between the members of a social group are and how homogeneous are their properties regarding the overall structure. The highest the cohesion of a group is then the minimal becomes the size of the cliques (subgroups) contained and formatted within the group.

According to Friendklin([Frienklin, 1984]) cohesion is a factor of influence of the individual by the group standards. A member of a group is highly connected with the group if and only if he/she accepts and/or possesses the characteristics of the groups. Let us consider for instance a set of international students that form the class for a course. Cliques/subgroups may be formed with students that share the same language/ethnic characteristics. A factor that will may affect the cohesion of such a group is the homogeneity of cultural standards and communication capabilities such as communication using a common language.

Cohesive subgroups pose an important research question regarding the emergence of social forces that influence their formation as well as the definitions of borderlines between the mem-

bers of the subgroup and the other members of the social structure. This comes with important connection to our previously formulated research question since the formalization of how well connected a member of a group is to the others is characterized by vague propositions where traditional approaches tend to quantify with means of poor expressiveness. Whenever this type of the social tie is considered several aspects of how strong this relation is have to be taken into account as well as their implications in different domains.

## 2.4 Social Network Analysis

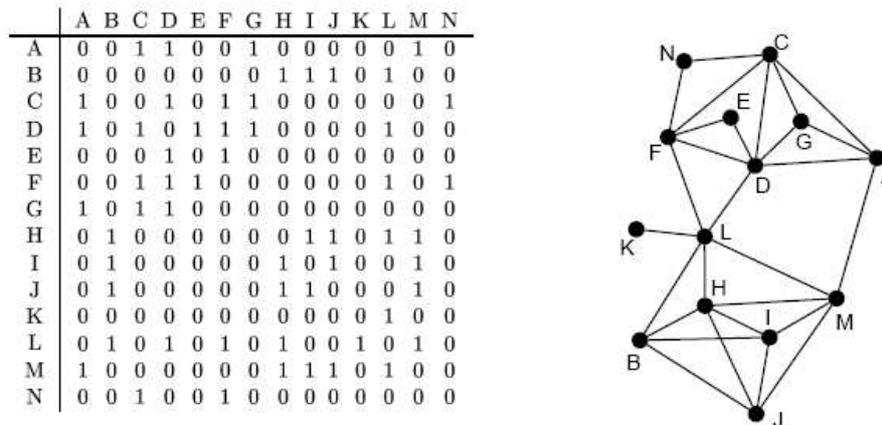
Social Network Analysis (SNA) is the study of sociometric models using formalizations from Graph Theory ([Harary, 1969]) as well as algebraic and statistical (in particular probabilistic) methods. In common principle SNA tries to unravel patterns of relationships that play a major role in the behavior of each individual thus becoming an important factor of the overall network activity. This can be observed clearly in situations known as the network effect where the decision of an individual comes as a consequence of each relationship (tie) with another member of the network. Usually this can be seen in political scenarios where one voter affects the decision of another, spreading also the notion that this opinion is the correct one.

Further foundations of Social Network Analysis and Sociometry in general can be found in the “gestalt” tradition in cognitive psychology ([Kohler, 1947]) where elaborations of studies regarding the structure of organized patterns of thoughts and perceptions gave the basis for extensions to group organization and behavior ([Leavitt, 1951],[Bavelas, 1948]).

Social network analysis and Sociometric studies in general considers two major assumptions when presenting results:

- That the relationship of individuals corresponds accurately to the real context. Subconscious or illicit relationships are not represented or either are not subject of the sociometric study.
- Group size is the optimal e.g. it includes actors that influence directly at least another member of the network or the social group that is examined.

SNA communicates its studies with the use of *sociograms*, first introduced by Jacob Moreno ([Moreno, 1946]). Among others a key usage of a sociogram is to communicate the principle of the sociometric star who is the social entity that receives the most choices (connections) in a social group thus dominates over the others.



**Figure 2.3:** A bidirectional sociogram along with its adjacency matrix

### 2.4.1 Formal Representations

In SNA a network is formalized as a graph  $G := (V, E)$  consisting of a set of vertices  $V = (v_1, v_2, v_3, \dots, v_n)$  that represents the social entities and a set of edges  $E = (e_{11}, e_{12}, e_{21}, \dots, e_{ij})$  where  $e_{ij}$  represents the adjacent connection between the nodes  $i$  and  $j$ . The network structure can be represented as a symmetric matrix (adjacency matrix) in which the nodes are listed in both axes and a boolean value is assigned to the  $e_{ij}$  which depicts the ties between the social entities represented in  $V$ . Figure (2.3) shows a sociogram along with the adjacency matrix. Depending on the type of the social network the adjacent connection may be:

- **Edge:** Indicates a bidirectional connection. In that case the graph is in the general form:  $G := (V, E)$
- **Signed Edge:** Indicates a bidirectional connection which is assigned with a value. In that case the graph is in the general form  $G := (V, E, E_d)$  where the set  $E_d$  is the value set for the mapping  $E \Rightarrow E_d$
- **Arc:** Indicates a directional connection. The formalization considers the general graph where we have  $G := (V, E, E_a)$  denoting that the connection is a directional
- **Signed Arc:** Indicates a directional connection where the arc is assigned with a value as to have  $G := (V, E, E_a, E_d)$  where  $E_a$  and  $E_d$  is the directed connection and value set respectively.

Furthermore depending on the group size a relation can be either dichotomous, trichotomous or contained in a subgroup.

A **dichotomous** relation is a bivalence type of relational tie where the valued set  $E_d$  is mapped to  $[0, 1]$ . Dichotomous relationships form units called dyads that are used to study indirect properties between two individuals.

A **trichotomous** or triad is a container of at most 9 ( $1 \bullet 2 \bullet 3$ ) dyads that can be either signed or unsigned.

A **subgroup** is a container of triads where the members are interacting using a common flow or path. A subgroup is not always a clique since the property is only topological and thus it doesn't depict any cohesion from the relational ties formed.

## 2.4.2 Design of Sociometric Studies

When conducting a social network study one might consider two critical points to the collection of data and the definition of boundaries to which this data will correspond.

### 2.4.2.1 Data Collection

In data collection the researcher needs to define the data that is going to be gathered and analyzed during the research. On this step one might define:

**Unit of Observation** The unit of observation is the entity upon which the analysis is focused. Depending on the context and the scale of the research a unit of observation can be people in different contexts (students, doctors etc) as well as larger entities such as nations (For example see the country trade network analysis in ([Wasserman et al., 1994])

**Scale** Depending on the kind of observation SNA may be employed to address different levels of structure in connection with the unit of observation. These levels may be:

- An actor: A single entity
- A dyad/triad: A dichotomous or trichotomous relation
- A clique/subgroup: An embedded structure in an another network
- A network: A large network member of a system (for instance a network of protein structure inside an organization).

**Relational Quantification** Relational quantification employs the modeling of the relational tie between the units of observation that are subject to the study and have been defined in the previous step. As denoted in (2.4.1) a relation is represented by an edge or an arrow

denoting whether this connection is directional or bidirectional. Depending on whether this relation contains a value or not the relational tie can be dichotomous or valued. The definition of relational tie is important for the level of expressiveness into which the social network is going to be modeled. For instance in non-directional relationships is not feasible to employ questions regarding roles and activity because we don't have information about the direction to which the relation is depicted.

In dichotomous connections on the other hand is not possible to quantify the strength or the weakness of a relational tie since the representation in the adjacency matrix is in a boolean form. If the relation is valued then the sociomatrix is accompanied by a vector denoting the values that depict the relational tie.

#### **2.4.2.2 Boundary Specification**

Another important step that the researcher has to define is the boundary to which the relational ties will be represented. Boundary specification in a network study is similar to the definition of the population in statistical analysis whether the extend to which the measurement and network definition will be carried needs to be addressed. However unlike in the case of population definition in statistics where the population members are assumed to be independent in network studies the axiom is that the entities are interdependent.

Depending also to the focus of the study a researcher might want to employ transitive properties between the members of the population. Lauman ([Laumann et al., 1983]) approaches the problem of boundary specification using two perspectives namely the "realist" and "nominalist". In the "realist" approach the researcher uses the actors perceiveness of the boundary while the nominalist employs the network properties that are subject to the individual such as sex, age, area etc.

Another important aspect on the boundary specification is the sampling of the entities that are going to be represented in the network. Sampling might be needed in cases where the acquisition of data is not possible or the population is too large. As it is with the definition of the population network, sampling is dependent upon the properties and in particular the connections, of the network entities. Knowing the properties of the network entities such as the degree (see 2.5.1.1) one might filter the network based on the properties of the network entities. Network sampling must not be confused with the definition of subgroups whereas the sampling employs filtering of entities based on their general properties where subgroup definition is dependent upon the cohesion of its members.

### 2.4.3 Applications of Social Network Analysis

Since its SNA has been emerged as a key tool in social research. As aforementioned the general goal is to unravel patterns of social communications and relations in every aspect eg. political exchanges, collaboration, trading etc. Furthermore contextualized interpretations of those relationships can be applied in cases such as:

**Epidemic Modeling:** Epidemic modeling ([Daley and Gani, 2001]) is the study of propagation models of viruses that are used to test ways of immunizing a social group from a virus thread (eg. through the use of simulation models - see for example ([Brouwers, 2005])). Along with the propagation characteristics a directed sociogram (see 2.4.1) is needed in order to examine the infection probabilities of each person depending on their spatiotemporal characteristics. Adoption and diffusion of innovation is a generic case of epidemic modelling with however less complexity since parallel stages such as vaccination are not present.

**Small World Phenomena:** Small world phenomena such as the infamous “six degrees of separation” examine the diameter of transitive relations in a network and the applications in the study of group communication such as group density and influence. Theoretical propositions of the small world have been formalized in 1970’s by Manfred Kochen (IBM) and Ithiel de Sola Pool (MIT) ([Sola Pool and Kochen, 1971]) who set out to prove the theory mathematically. The hypothesis is formalized as follows:

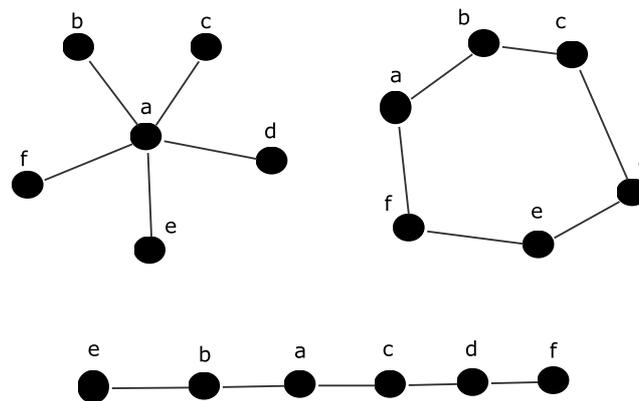
Given a set of  $N$  people, what is the probability that each member of  $N$  is connected to another member via  $k_1, k_2, k_3 \dots k_n$  links ?

In 1967 the American sociologist Stanley Milgram devised a new way to test the theory, which he called "the small-world problem". He randomly selected people in the mid-West to send packages to a stranger located in Massachusetts. The senders knew the recipient’s name, occupation, and general location. They were instructed to send the package to a person they knew on a last-name basis who they thought was most likely, out of all their friends, to know the target personally. That person would do the same, and so on, until the package was personally delivered to its target recipient.

The outcome of Milgram’s experiment (which later has become a common concept) was that the average of connections was six which despite the fact that people participate in a huge number of social groups with a big density the distance is very short

Type of Study	Literature
Diffusion of Scientific ideas	[Mullins, 1973],[Doreian and Fararo, 1985]
Economic Markets	[Berkowitz, 1988],[Leifer and White, 1987]
Group Problem Solving and Distributed Cognition	[Bavelas, 1948],[Romney and Faust, 1982]
Adoption and Diffusion of Innovation	[Rogers, 1979],[Colamen et al., 1957]

**Table 2.1:** Contextualized SNA studies in the Social Sciences



**Figure 2.4:** The Three basic types of rankings in a network: Star, Circle and Line

([Milgram, 1967]). Although Milgram’s experiment has been proved to have some truth to our everyday life it is not the same for the web. Albert and Barabasi in a series of experiments measured that the average distance in the web is about 19 intermediaries or clicks ([Albert et al., 1999]).

Furthermore in the social sciences there have been many examples of applications of SNA in various contexts. Table 2.1 summarizes those contextualized interpretations of SNA studies.

Depending on the context a network study may employ different kinds of measurements. In this thesis we are interested on indices that measure ranking and trustworthiness in a social entity and its properties. Those indices are based on graph theoretic properties of the social network and are presented in the section that follows.

## 2.5 Indices of Social Ranking

In a society one recognizes several strata that characterize their members with a kind of implicit rank which is known in the social science literature as “status” ([Katz, 1953]). Usually status

denotes power expressed in different contexts such as political or economical. The most basic theoretical implication of status is the *availability of choices the entity receives in the network which gives the entity the advantage of negotiation over the others*. Depending on the topology of the network status can be attributed to several nodes of the sociogram (see figure 2.5).

In SNA studies, status is depicted upon the generalization of the location of the actor in the network. In particular status is addressed by how strategic is the position of that entity in a network (e.g how it affects the position of the others). Theoretical aspects of status were defined by Moreno and Jennings ([Moreno and Jennings, 1945]) as the instances of the sociometric “star” and “isolate”.

Quantifications of status employ techniques of graph theory such as the centrality index ([Sabidussi, 1966]) which have been adjusted to the various representations of ties in a network. In our case we summarize the two most noteworthy measures of an actor in a **directed** network which are the “Prestige” and “Centrality”.

### 2.5.1 Prestige

A prestige measure is a direct representation of status which employs the non-reciprocal connections/choices provided to that entity along with the influence that this entity might provide to the neighboring entities.

#### 2.5.1.1 Indegree

The simplest measure of prestige is the indegree index of that entity which is normalized as follows: Considering a graph  $G := (V, E_A)$  and  $E_A$  the set of the directed connections  $E = (e_{1,1}, e_{1,2}, \dots, e_{i,j})$  between the members of the set  $V$  then the indegree  $K_i^{in}$  of a vertex  $V_i$  is the sum of the incoming connections to that vertex.

$$K_i^{in} = \sum_{j=1} e_{j,i} \quad (2.1)$$

#### 2.5.1.2 Degree of Influence

However the indegree index of a vertex makes sense only on cases where a directional relationship is available (the connection is non-reciprocal) and this cannot be applied in the study of non-directed networks. In that case the prestige is computed by the influence domain of the vertex. For a non directed graph  $G := (V, E, \bar{E})$  the influence domain of a vertex  $V_i$  is the number or

proportion of all other vertices which are connected by a path to that particular vertex.

$$\bar{d}_i = \frac{1}{N-1} \sum_{j=1} \bar{e}_{j,i} \quad (2.2)$$

whether  $\bar{E}$  represents the set of paths between the vertices  $V_i$  and  $V_j$  and  $N-1$  is the summation of all available nodes in the graph  $G$  (The total number of node  $N = |V|$  minus the node that is subject to the metric).

### 2.5.1.3 Proximity Prestige

A combination of the above two metrics of prestige is known as the ‘‘Proximity Prestige’’ ( $PP_i$ ) of vertex  $V_i$  encompasses the normalization of the indegree of the vertex by its degree of influence such as:

$$PP_i = \frac{K_i^{in}}{\bar{d}_i} \quad (2.3)$$

A special case of prestige is used in the HITS algorithm (see 3.3.1) to compute the relative hubness and authoritativeness of a web page in the web.

## 2.5.2 Centrality

Unlike prestige measures which rely mainly on directional relations of the entities the centrality index of a graph can be calculated in various ways taking also into account the non-directional connections of the vertex which is examined. The most noteworthy measures of centrality are classified by the degree of analysis they employ in the graph.

### 2.5.2.1 Actor Degree Centrality

Actor degree centrality is the normalized index of the degree of an actor divided by the maximum number of vertices that exist in a network. Considering a graph  $G := (V, E)$  with  $n$  vertices then the actor degree centrality  $C_d(n_i)$  will be:

$$C_d(n_i) = \frac{d(n_i)}{n-1} \quad (2.4)$$

Where  $n-1$  is the remaining nodes in the graph  $G$ . Actor degree centrality is often interpreted in the literature as the ‘‘ego density’’ ([Burt, 1982]) of an actor since it evaluates the importance of the actor based on the ties that connect him/her to the other members of the network. The highest the actor degree centrality is, the most prominent this person is in a network since an actor with a high degree can potentially directly influence the others.

### 2.5.2.2 Closeness or Distance Centrality

Another measure of centrality the *closeness centrality*, considers the “geodesic distance” of a node in a network. For two vertices a geodesic is defined as the length of the shortest path between them. For a graph  $G := (V, E)$  the closeness centrality  $C_{c(V_i)}$  of a vertex  $V_i$  is the sum of geodesic distances between that vertex and all the other vertices in the network.

$$C_{c(V_i)} = \frac{n - 1}{\sum_{j=1}^{n-1} d(v_i, v_j)} \quad (2.5)$$

Where the function  $d(v_i, v_j)$  calculates the length of the shortest path between the vertices  $i$  and  $j$  and  $n - 1$  is the number of all other vertices in the network. The closeness centrality can be interpreted as a measurement of the influence of a vertex in a graph: the higher its value, the easiest it is for that vertex to spread information into that network. Distance Centrality can be valuable in networks where the actors possess transitive properties that are spread through direct connections such as the case of epidemic modeling.

### 2.5.2.3 Betweenness Centrality

*Betweenness* is the most celebrated measure of centrality since not only measures the prominence of a node based on the position or the activity but also the influence of the node in information or activity passed to other nodes. Considering a graph  $G := (V, E)$  with  $n$  vertices, the betweenness  $C_B(v)$  for vertex  $v \in V$  is:

$$C_B(v) = \frac{\sum_{s \neq v \neq t \in V} \sigma_{st}(v)}{(n - 1)(n - 2)}$$

Where  $\sigma_{st}(v) = 1$  if and only if the shortest path from  $s$  to  $t$  passes through  $v$  and 0 otherwise. Betweenness can be the basis to interpret roles such as the “gatekeeper” or the “broker” which are studied extensively in communication networks. A vertex is considered as a “gatekeeper” if its betweenness and indegree is relatively high. The “broker” is a vertex which has relatively high outdegree and betweenness.

This chapter gave to the reader an insight from the field of social network analysis and the measures employed to analyze the prominence and the rank of a social entity based on the relational ties. In the next chapter we present some measures from information retrieval and in particular search engine research which examine a different kind of networks: that of web pages and their hyperlinked structure.

## Chapter 3

# Ranking in Information Retrieval

### 3.1 Background

Information Retrieval (IR) ([Frakes and Baeza-Yates, 1992]) employs the construction and the application of models that address information inquiries submitted by the users. The evolution of WWW and storage technology as well as advances of computing have increased the ability of retrieving relevant information over an enormous amount of information resources. According to industry reports ([Prete et al., 2003]) internet search engines are the most important part of the web and their usage accounts for the 55% of the activity of internet users. As the information on the web increases one primary concern in search engine research is not only to find the most relevant result in a query but also to evaluate the “precision” of this query by using several importance metrics, so that the query with the highest precision will be ranked in the top.

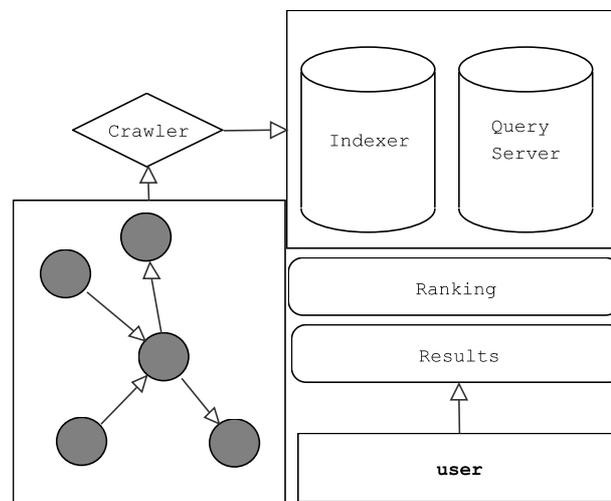
Search engines like Google or Yahoo! use such importance metrics to filter results and rank them according to the relevance in the query term as well as the popularity that this information resource might have in the hyperlinked context. Although this context can be used to extract importance, ranking models are always subject to human evaluation whereas cases such as the relevance of the resource is an “a-posteriori” analysis that must be carried due to the complexity of evaluating “meaning” as a cognitive task.

While several methods and techniques of information retrieval have been developed such as lexical analysis and text retrieval our focus is on ranking models that are applicable in hypermedia context. Ranking models in the web employ the same formalization as social network indices of ranking with the primary difference been in the way that connections are directional (eg. a page links to another) and the pairs are either dichotomus or signed. For instance a page might be linked to another (referrer) but is not necessary that the linked page

will link back to its referrer.

## 3.2 The Web as a graph

Several studies have been done to model the web as a large graph which consists of nodes (pages) and directed edges - arcs (hyperlinks) between them. As aforementioned the formalization is quite similar to a social network however studies that employ the density of the graph are characterized by high complexity since the web is a complex system ([Newman, 2003]) into which resources are added in a high rate, therefore making accurate metrics of the web is a rather complex task. Graph properties of the web ([Kumar et al., 2000]) incorporate the use of power law's to explain the link structure as well as using properties of random graphs based on the work of Erdos and Renyi ([Erdos and Renyi, 1960]). Most of the work in modeling the web as graph employs models from random graph theory ([Bollobas, 2001]) applied widely in cases such as the planning of crawling strategies where a web crawler needs to traverse through the web graph and index the information in the indexer, to be used later by the query server. (See the figure 3.2)



**Figure 3.1:** Typical architecture of a web search engine as a user-query interface

A formalization using elements from the theory of random graphs can be as follows: Let us consider a directed large graph  $G_U := (V, E)$ . For this graph model we consider the following stochastic processes :

- $Vr_c$  is the vertex creation process

- $Er_c$  is the edge creation process
- $Vr_d$  is the vertex deletion process
- $Er_d$  is the edge deletion process

$Vr_c$  and  $Er_c$  are independent processes which means that when a node is created it doesn't imply that an edge will also be created whether  $Vr_d$  and  $Er_d$  are interdependent which yields that when a vertex is deleted then all the edges (hyperlinks) going to and coming from the deleted page (vertex) are also deleted.

Page popularity is measured by employing various models, however in this thesis we focus on the citation or back-link models that measure page importance since (a) these metrics are the most popular/implemented in the search engine industry and (b) our goal is to bias those metrics by employing the social context of their creators in the consideration of both hyperlink and social layers.

### 3.3 Citation and Back-link Models

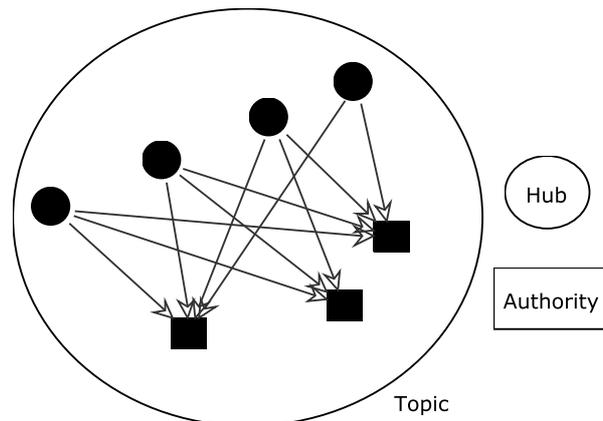
Citation models consider the indegree characteristics of a web page to measure its popularity. Early usage of such models has its roots to scientometrics where the analysis of the graph of scientific citations has been subject to major observations in the beginning of 20th century. Most notable among them is Lotka's Law ([Lotka, 1926]) which implies that the number of contributing authors in scientific productions making  $n$  contributions is about

$$f(t) = \frac{1}{n^a} \quad (3.1)$$

of those making one contribution, where  $a$  is often nearly 2. Lotka's law can be also seen as a power law or "inverse square law". The word law in that context implies the statistical distribution that is prescribed by the above equation. Power laws characterize mostly the degree models of the web since they try to evaluate strategies deployed by web crawlers using properties of their outer and indegree distributions ([Faloutsos et al., 1999]).

#### 3.3.1 Hubs and Authorities

The concepts of hubness and authoritativeness of a node (web page) are based on the intuition that the most important pages will be cited by pages that are also important and will cite pages with particular importance. Graph theoretic formalizations of the above concept can be described as:



**Figure 3.2:** Hubs and Authorities contained in a topic

- The **hubness** of a node is the (outer)degree to which a node links to other important authorities
- The **authoritativeness** of a node is the (inner)degree to which a node is pointed to by important hubs

Resolving the hubness and authoritativeness of a node requires is a mutual recursion process that needs to be addressed algorithmically.

### 3.3.1.1 The HITS algorithm

Original work on modeling web structure considering the hubness and authoritativeness of a particular node has been carried out by Kleinberg ([Kleinberg, 1999]) which resulted the Hypertext Induced Topic Selection (HITS) algorithm. In HITS hubness and authoritativeness is defined by two general operations:

- $I^{op}$ : Where a good authority is pointed by many hubs
- $O^{op}$  Where a good hub is pointed by many authorities

The above operations can be quantified as:  $x = I^{op}(y)$  and  $y = I^{op}(x)$  where:

$$X = (x_1, x_2, \dots, x_n)^T \text{ and } Y = (y_1, y_2, \dots, y_n)^T \quad (3.2)$$

represent the authority and hub score of each page respectively.

### 3.3.2 The PageRank Citation Ranking

The intuition behind PageRank is that it uses information which is external to the nodes themselves, their back-links, which provide a kind of peer review ([Page et al., 1998]), in order to identify some relevant importance about their position inside the network. Metaphorically PageRank can be seen as a result of a ballot among the pages that are contained in a linked set. A vote of page  $A$  to page  $B$  is a hyperlink originating from  $A$  to  $B$ . Furthermore the page that casts the vote is also subject to the evaluation of the others resulting in an iterative process that spans all the members of the page set.

Mathematically PageRank can be modeled as follows: We consider a graph  $G := (V, E)$ . For the node  $k_i$  belonging to the set  $V$  such as  $k \in G$  then the PageRank of the  $k_i$  is

$$PageRank(k_i) = \frac{q}{N} + (1 - q) \sum_{k_j \in M(k_i)} \frac{PageRank(k_j)}{L(k_j)} \quad (3.3)$$

where

- $k_1, k_2 \dots k_N$  are the vertices under consideration
- $M(k_i)$  is the set of vertices that link to  $k_i$
- $L(k_j)$  is the number of links coming from node  $k_j$
- $N$  is the total number of vertices (pages) contained in the graph.

The variable  $q$  is a dumping factor which is used to initialize the calculation process.

Although applications and variances of the PAGERANK algorithm have been presented in the literature since 1970's and in particular into the measurement of impact factors of scientific journals ([Pinski and Narin, 1976]), early deployments of the PageRank were not commercially available because of the complexity that characterizes the iterative process of calculating the PageRank for each page of the very large graph of the web. Google founders ([Brin and Page, 1998]) employed apart from the usage of the algorithm the deployment of a distributed computing system based on clusters formed from inexpensive Linux based servers that could result a trustworthy calculation of the PageRank of each pages thus bias properly the results obtained from the query servers making the pages with the highest popularity to appear in the first positions.

On the other hand the importance of search engine ranking in e-commerce business models as a directional cause of traffic into the website/e-shop has resulted the deployment of various methods to bias the calculation of PageRank to be higher for pages that deserve to have less PageRank. Those methods known as "spamdexing" employ the implementation of link farms that

point to the pages that try to be listed in the top results. TrustRank ([Gyongyi et al., 2004]), a variance of PageRank, targets to address the “spamdexing” by penalizing the pages that are pointed by the link farms.

PageRank and HITS conceive the web as a directed dichotomous graph where each connection is represented by a hyperlink denoting only the destination and the source of the connection. The second generation web or the “semantic” web is a web where the dichotomous tie is replaced by a valued one resulting the creation of a signed directed graph where more sophisticated metrics may be employed to result higher trustworthiness in ranking.

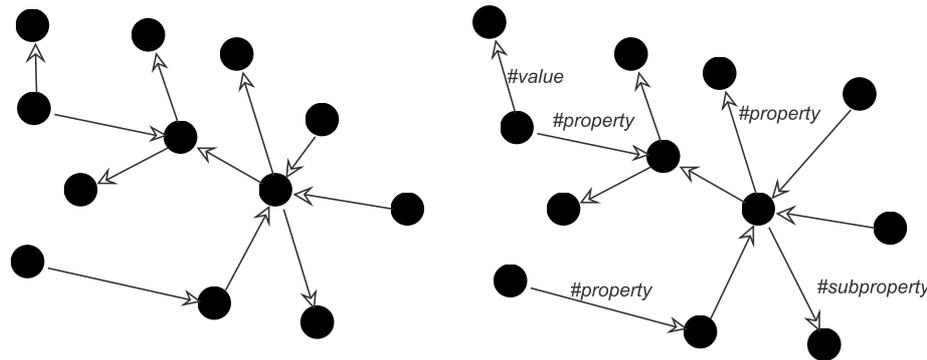
### 3.4 Semantic Web

Undoubtedly the web as a mean of communication has transformed the way people interact and express their thus has resulted an information ecosystem where information resources and the entities they represent (such as authors, groups or organizations) are encoded and linked in a way that targets direct human processing and interaction. The aforementioned ranking models use tools such us lexical analysis and graph analysis to extract the link structure thus evaluate the trustworthiness of the node based on citation models that are expressed through relatively “poor” means. For example the PageRank algorithm considers as a back-link to a resource a reference to that resource from another through the standard syntax of hyperlinks as in the form:

```
<a href='reference'>Reference to the resource </a>
```

However the current expression of that syntax cannot encapsulate the meaning of the tie under several contexts. In order to see a more practical issue let us revisit the paradigm of scholar evaluation. An information resource that represents an article gets several pointers from other information resources that also represent articles. However in the context of scientific citation and referencing, one might want to know the meaning that this connection might have. Does the research want to cite it as a negative example or as a classic work in the field ? .The current expressiveness of the web cannot address such important properties although work on extending the expressiveness of hyperlinks has been carried out by W3C with the proposal of XLink([DeRose et al., 2001]).

The Semantic Web ([Berners-Lee et al., 2001]) is a vision of creating a machine processable web where information resources will be interconnected using machine processable expressions denoting the “semantics’ of these connections thus providing a base for intelligent agents and



**Figure 3.3:** Network of pages in the current and the semantic web

other software to apply reasoning techniques to easily extract relevant information from the web. As the original designer of WWW, Tim Berners-Lee, states:

*“...new forms of collaboration will emerge, allowing people and machines to work together[Berners-Lee et al., 2001]”*

As can be seen in the figure 3.4 in the semantic web an information resource is characterized by meaningful connections that denote properties of this resource and values or literals that are attached to this property. Let us consider the following proposition:

*Nikos is a Masters Students that lives in Stockholm and takes a course in Royal Institute of Technology which is a university located in Sweden.*

The above proposition although seems complete it lacks information such as the location of the university and the country into which Nikos leaves therefore when a crawler accesses Nikos’ page to index information about Nikos will not be able to capture all the information correctly since its fairly incompletely expressed. To address that case in the semantic web the information will be encoded with context independent machine processable language such as expressions based on the resource description framework (RDF)([Klyne and Carroll, 2002]) or more primitive ones such as N3 predicates. A N3 expression of the following proposition can be done follows:

```
:nikos :studies [ :Royal Institute of Technology ]. :lives [ : Stockholm ].
:Royal Institute of Technology :located [ :Sweden].
```

In that case the crawler may use inference to obtain knowledge that Nikos is living in Sweden and that the Royal Institute of Technology is a university located in Stockholm.

According to Gho and Garcia Molina ([Cho and Garcia-Molina, 2000]) inference abilities of the crawler will contribute to decrease the overhead of the cite crawling process and the index update time will be shorter since the crawling of the graph will be done quickly and more effectively. One particular model of expressiveness in the Semantic Web is the way authors of information resources contribute to the identification of those information resources as their productions as well as to the contextual (social) connections that are depicted on the linking of these resources.

### 3.4.1 The FOAF Vocabulary

The Friend-of-a-Friend vocabulary ([Brickley and Miller, 2005]) is an expressive vocabulary set which syntax is based on RDF technology that is gaining popularity nowadays as it is used to express the connections between social entities in the web along with their hypertextual properties such as their homepages or the emails. In a best case scenario the author of a web page (information resource) will attach his FOAF profile in the resource in order to make it identifiable as an own production by the visitors of that page. This can be observed clearly in cases such as Blogs where the information resource represents the person that expresses his/her views through the blog. Furthermore connection between blogs represents also a kind of a directional dichotomous tie between the authors of those blogs. In FOAF standard RDF syntax is used to describe the relations between various acquaintances (relations) of the person described by the FOAF profile. This relation is depicted in the *<foaf:knows>* predicate which denotes that the person who has a description of *<foaf:knows>* in his profile for another person, has a social connection with that person as well. For example a FOAF profile for the author of this thesis and the connection he has with his supervisor can be described by the following fragment of RDF code.

```
<rdf:RDF xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
xmlns:rdfs="http://www.w3.org/2000/01/rdf-schema#"
xmlns:foaf="http://xmlns.com/foaf/0.1/">
<foaf:PersonalProfileDocument rdf:about="">
<foaf:primaryTopic rdf:nodeID="me" />
</foaf:PersonalProfileDocument>
<foaf:Person rdf:nodeID="me">
<foaf:name>Nikolaos Korfiatis</foaf:name>
<foaf:givenname>Nikolaos</foaf:givenname>
```

```
<foaf:family_name>Korfiatis</foaf:family_name>
<foaf:nick>nkorf</foaf:nick>
<foaf:mbox rdf:resource="mailto:n-korf@nada.kth.se" />
<foaf:homepage rdf:resource="http://www.nada.kth.se/~n-korf" />
<foaf:knows>
<foaf:Person>
<foaf:name>Ambjorn Naeve</foaf:name>
<foaf:mbox rdf:resource="mailto:amb@nada.kth.se" />
<rdfs:seeAlso
rdf:resource="http://www.nada.kth.se/~amb" />
</foaf:Person>
</foaf:knows>
</foaf:Person>
</rdf:RDF>
```

Research on descriptions of social relations in the semantic web is an undergoing effort which has been initiated lately<sup>1</sup> to address the various concerns and sociological implications for the expressiveness of social connections in the web. One particular issue is that although the FOAF vocabulary (see the table 3.1) has a set of properties for the description of several kinds of relationships such as genealogic (ancestorof, decenstorof) the *<foaf:knows>* property is the most common relation that is expressed in a publicly available FOAF profile. According to the general discussion in the FOAF project the reason for this is that many users prefer not to express their strength of social connections publicly than to do it with a general way which the *<foaf:knows>* property implies.

However description of social relations available on the web are also a subject of privacy research since publications about the social ties of person is something that needs to be carefully examined before done because of the various methods of the so called “social engineering” (in the relm of information technology)<sup>2</sup> and also forgery and several other kinds of deceptions (e.g. phishing).

The construction and the availability of FOAF profiles and their social properties in the Semantic Web gives input to processes of determination of the social ranking of the authors of information resources along with the hypertextual context obtained from the link structure. As noted in Chapter 1 the aim of this thesis is the examination and construction of a measure that

---

<sup>1</sup>See the FOAF Project homepage: <http://www.foaf-project.org>

<sup>2</sup>see [http://www.webopedia.com/TERM/S/social\\_engineering.html](http://www.webopedia.com/TERM/S/social_engineering.html) for a definition.

Vocabulary Element	Description	Type of Relation
foaf:knows	links foaf:persons	<b>direct</b>
foaf:member	Provides affiliation/membership (relates an entity with a social group)	indirect
foaf:maker	Indicates authorship (relates an information resource with it's creator)	indirect
foaf:based_near	Indicates a spatial affiliation of a social entity	indirect
foaf:currentproject	Indicates a temporal affiliation of a social entity with a project or an activity	indirect

**Table 3.1:** Basic Elements of the FOAF Vocabulary and their type relational tie

will evaluate social connections as a bias to the ranking obtained by the link structure in order to result a more trustworthy metric of web ranking.

However a dichotomous relation lacks expressiveness when comes to denote such a complex property therefore we deploy the use of different methods of expressiveness which are analyzed for the social context on the following chapter.

## Chapter 4

# A fuzzy approach on the representation of social ties

Fuzzy logic is a term widely known in the field of control engineering and automation because of numerous implementations on home appliances and control systems. It was introduced by Lofty Zadeh ([Zadeh, 1969]), as a calculus for vague meanings where traditional boolean logic could not address. The basic concept behind fuzzy logic is the concept of partial truth of a statement where a proposition is being represented by degrees of truth. Traditional logic cannot handle such vague propositions since it will violate the “principle of bivalence” in which is stated that a proposition cannot have membership to both separate crisps denoted by a boolean value of membership.

To our thesis Fuzzy Set theory came as an important input to the degree of representation of social relationships employed by the structure of the social network. As we are going to argue later social relationships and their categorization to property sets denoting different association typologies are characterized by a relatively large set of properties that sign those relationships therefore any crisp set formalization does not possess the expressiveness to capture the input and the implications of those relationships to the phenomena that are studied and observed in the social structure.

### 4.1 Traditional Representations

Several sociologists have studied and defined aspects of social relationships that incorporate both directional and non-directional ties between the social entities connected. According to Berger and Luckman ([Berger and Luckmann, 1967]) the type of the relationship can be affected

Traditional Representation	Fuzzy (Linguistic) Expression
Friend	A good friend, A colleague, A person i know
Good	Very Good, Quite Good
Bad	Very Bad, Quite bad

**Table 4.1:** Traditional and Fuzzy(Linguistic) Expressions of Directional Relations in a Social Network

by two major factors namely the *context* and the *purpose*.

**Context** represents the social setting under a social tie is being formed. Depending on how the contextual surrounding affects the actor, different instance of a relational tie might occur. For instance let us consider two members of a family that have a genealogic relation. Depending on the family and the societal structures such as cultural influences the relation might be weak or strong.

On the other hand **purpose** is visible mostly in directional relations since that kind of relational tie represents flow from one entity to another. Depending on the type of relation purpose can be seen as information, transaction, flow of resources etc. As we already saw in section (2.5.2.3) the betweenness of the actor that receives the flow is a metric of his/her importance in the network structure.

However context and purpose are actor characteristics and although might influence the formation, do not directly relate with the social ties between two or more actors. In that case we may extend our focus to interdependent properties as to involve a richer level of expressiveness upon the vague propositions that characterize this properties.

A tie depending on the context and the purpose may have several instances. For example in a case of a directional relation a social tie is used as a way of initiating a communication channel from the one entity (source) to the other (destination). If that channel is not often used then this communication is weak since the flow from the source to the destination is not effective enough, disregarding the entity's attributed values (purpose and context). This can be seen as a metaphor for a bottleneck. A weak tie is a bottleneck in networks where a flow must be communicated (eg. a transaction flow). The measuring of that capability provides us an intuition to consider a basic parameter to our model of social tie, that of strength.

### 4.1.1 Strength

Although original work in sociometry questioned the position of actors in a network and their connectedness in a group, the extend to which this connectedness of a particular actor with the others is defined was not fully investigated empirically. To address that issue the concept of the particular “strength” of a tie was introduced by Granovetter ([Granovetter, 1982]) which models empirically the concept of connectedness in a network topology. Furthermore network studies have questioned the ability to model not only the formal representations of strength but also the affective ([Krackhardt, 1992]) such as notions of "philos" and cognitive “passions” similar to those expressed by Heider in balance theory with the difference that in this case examination is not focused on the topological implications of affective ties and not the local.

The representation of strength till now is being done formally using network measures that hold thresholds in order to categorize different scales of that measure. However the issue arises on *how to define the borderline of the measuring levels that form that scale.*

A weak tie as Granovetter denotes is more likely to expose it’s weakness in an isolation from the other entities of the structure. However it often reestablishes itself with another destination. The possibility of this to happen depends on context dependent properties such as affiliation with a third party or a common activity. This provides input to the concept of a particular *evidence* to be considered as a parameter affecting the relevant trustworthiness of the tie.

### 4.1.2 Evidence

Often in social relationships actors use the presence of a third entity to establish a relationship. The relationship with that entity can be an affiliation or a common activity. Nonetheless the presence of that particular entity provides a common ground for the establishment of that relationship. In that case evidence runs as a catalyst for the formation of the relational tie. However relational ties that have are characterized by an *evidence* are not formed as a consequence of exposure of actor attributes but rather as a biased formation depending on the context.

## 4.2 Fuzzy representations of Strength and Evidence

In fuzzy logic a proposition is represented by a fuzzy variable, which is a variable representing the linguistic expression under which that variable is evaluated. For instance if we want to represent the friendship between two persons, we can express it linguistically by saying “friends” or “close friends” or even “no friends”. These linguistic expressions employ the definition of scale whereas

the intermediaries between the top and the bottom may vary significantly. Let us consider the expression of an opinion regarding the usage of a product. One might say that “The Product is good” or “The Product is very good” or the “Product is ok”. Whereas it comes to evaluate and formally represent this expression the difficulty is on the position of the above in the range of the evaluation scale as well as the determination of the boundaries upon which the scale will span.

As already mentioned the basic intuition behind Fuzzy set theory is that several properties in our everyday surroundings, including our social ties, are vague therefore distinct boundaries provide fragmentations that might exclude some aspects of these cases. For instance one might say that “I’m strongly connected with my friends” but there are several kinds of vague propositions in this sentence such as those depicted in the table 4.1. The strength of the relational tie in a fuzzy expression is the linguistic expression under which an actor finds the best association with his/her cognitive model.

### 4.2.1 Representation of Sets

Formally in fuzzy set theory the strength  $S$  and the evidence  $Y$  that characterize a relational tie can be modeled as members of the fuzzy sets:

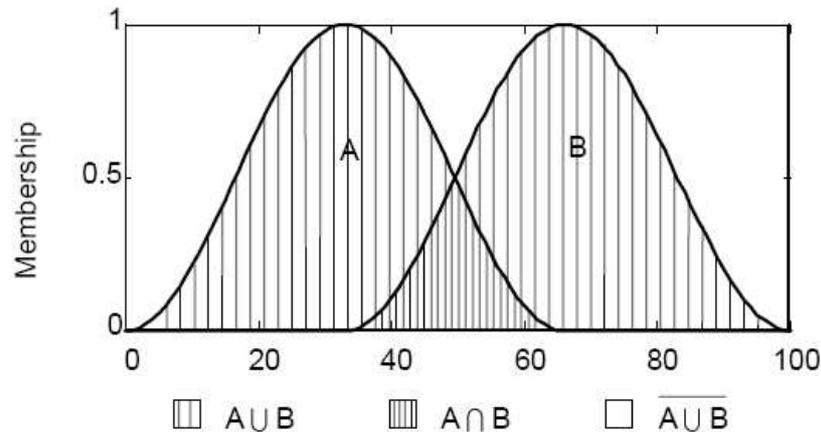
$$A = \{Weak, VeryWeak, NotSoWeak, MoreOrLessWeak, QuiteWeak\} \quad (4.1)$$

$$B = \{Strong, VeryStrong, NotSoStrong, MoreOrLessStrong, QuiteStrong\} \quad (4.2)$$

where  $A$  encompasses the well connected and  $B$  the less connected ties that are attributed with the variables  $S$  and  $Y$  as  $(S, Y) \in (A, B)$ . Figure 4.2.1 shows the graphical representation of the basic operations among these fuzzy sets. The intersection of  $A$  and  $B$   $A \cap B$  is the case of partial truth of the a relationship belonging both to the sets  $A$  and  $B$  (being either strong and weak) while  $\bar{A} \cup \bar{B}$  represents the case where membership in the set doesn’t exist which simply implies that a relationship is not established. The union of the two sets is defined in fuzzy set theory as the universe or discourse of the variables. For the above two sets ( $A$  and  $B$ ) we have:

$$U = A \cup B \quad (4.3)$$

such as  $S, Y \in U$  where the membership is subject to the standard operations provided to the sets (union, intersection and complement). As can be seen by the figure 4.2.1 memberships in the above sets is not bivalued but is provided by a membership function which represents the value range between the transition of the set  $A$  to  $B$ .



**Figure 4.1:** Basic Operations on the fuzzy sets  $A$  and  $B$  and the value of their membership function along with the transition from the one fuzzy set to the other

#### 4.2.2 Representation of Membership

The range of the universe of the linguistic variables denoting a relational tie is provided by the membership function  $\mu \rightarrow [0, 1]$  having its polarity on the values “weak” and “strong” respectively with a middle value (0.5) to represent the transition from  $A$  to  $B$ . Therefore for a variable  $x$  the degree of membership to the sets  $A$  and  $B$  will be given by the following:

$$A = (x, \mu_A(x)) \quad (4.4)$$

$$B = (x, \mu_B(x)) \quad (4.5)$$

As we will see on the implementation stage based on that definition we can evaluate membership of the variable  $x$  provide phases of transition inside the sets  $A$  or  $B$ . For instance we may wish to model the transition/boundary from being “Weak” to “Very Weak”. This output will have to defuzzified (quantified to a value that better represents the model) so it can be used as an input to a model of social relevance. Several methods of defuzzication exist for a particular task ([Yager, 1988]). In our thesis we selected an Aggregation Operator in order to have an overall picture of the attributed properties on the social network.

### 4.3 Aggregation

In order to provide a representative metric of strength and evidence in a relational tie we need to combine those two fuzzy variables using a technique which is called aggregation. Aggregation is a critical operation in systems that are affected by parameters with different implications thus provides an overall expression of the total input or output that accompanies the operation of a system. Mathematically aggregation in fuzzy systems is provided by an operator which has the property of reducing the set of relevant properties represented by values into a unique representative (input) number, to be used for further analysis. The need for aggregation operators stands in cases where the fuzzy variables are too disperse and need to be reduced for processing or as it happens in our case to be considered as an input to a non fuzzy determination. For our approach we chose the OWA operator which is the most common in the family of aggregation operators and freely available implementations of that operator are available for deployment.

#### 4.3.1 The Ordered Weighted Average Operator

The Ordered Weighted Average (OWA) is a well celebrated aggregation operator that was primarily used for the aggregation of scores contributed by multiple criteria. The distinct property of the operator is the generalization of the minimum and the maximum aspect of the criterion submitted to the operator. Formally the operator can be represented as follows:

$$OWA(x_1, x_2, \dots, x_n) = \sum_{j=1}^n W_j \cdot x_{\sigma_j} \quad (4.6)$$

where  $\sigma(j)$  is a permutation that orders the elements  $x_{\sigma(j)}$  as

$$x_{\sigma(1)} \leq x_{\sigma(2)} \leq x_{\sigma(3)} \dots \leq x_{\sigma(n)} \quad (4.7)$$

and having that the weight factor satisfies

$$w_j \geq 0 \quad (4.8)$$

$$\sum_{j=1}^n W_j = 1 \quad (4.9)$$

for all the  $i \in n$

### 4.4 Opinion as an aggregation function

Having formalized the above fuzzy sets and their basic operations for the non reciprocal relational tie of the entity  $n_1$  to the entity  $n_2$  we define an opinion of  $n_1$  to  $n_2$  as the Ordered

Weighted Average Aggregator of the fuzzy properties **Strength (S)** and **Evidence (Y)** as to have:

$$Opinion(n_1 \rightarrow n_2) = OWA(n_1 \rightarrow n_2) = (S_{1 \rightarrow 2} \bullet Y_{1 \rightarrow 2})_{n_1 \rightarrow 2} \quad (4.10)$$

In fact opinion represents the range of transition between the negative ( $A$ ) to the positive ( $B$ ) set of values depicted on the normalized inner-degree index that represents the signed directional/non-reciprocal connections between the nodes  $n_1$  and  $n_2$  of the social network.

Having formalized our model for opinion evaluation we provide an example implementation using first an algorithmic construction of the relevant index and then a high level implementation using the JAVA programming language.



## Chapter 5

# OpinionRank: A fuzzy adapted ranking model

### 5.1 Qualitative Exploration

Having previously discussed the fuzziness and the fuzzy set approach on modeling social ties we come back to the web retrieval aspects of ranking that as we have already discussed have their source to the link structure. However it is common sense that author credibility and trustworthiness play a significant role for the ranking of his/her productions which come in tight connection with their author. The algorithm proposition is based on two basic assumptions:

- Intuitively, the trustworthiness on the quality of an information resource is depicted to the degree of confidence we hold to their authors
- Pages authored or affiliated with people that hold a larger positive status should somewhat considered having better precision than other relevant pages who also satisfy the query requirements.

Interpretations of the above two propositions can easily be observed in the context of scholar research where one might wants to find which publication on that topic that satisfies the query is the most trustworthy. However in the case of web authorship the resulted network is a special case of social network where the pages are affiliated with their authors. In social network analysis this kind of social network is referenced as an affiliation network. In that network one might need to conduct two different types of associative rankings that will consider both contexts in order to extract the trustworthiness.

### 5.1.1 OpinionRank

We define OpinionRank as a ranking index containing the rankings of nodes obtained by PageRank biased with the opinion (OWA) obtained by the non reciprocal directed edge attributed properties strength and evidence. OpinionRank is a social layered version of the PageRank algorithm that is reconfigured to rank in the first positions the products that are both respected in the social contexts (through their affiliation with authors) and the link context. In fact OpinionRank is a normalization of the ranking index obtained by PageRank where the defuzization of the ties - using the OWA Opinion operator - is depicted on their final ranking as an input to the link context.

## 5.2 Implementation

Formally we represent the network that holds both contexts as  $G := \{D, S, E\}$  where:

$D$ : is the information resource subgraph that holds the web documents such as  $D \subseteq G$

$S$ : is the social entities subgraph that holds the affiliated authors such as  $S \subseteq G$

Both social and document layers are combined using an algorithmic invariant of the PageRank algorithm (see section 3.3.2). The algorithm is written as follows.

### 5.2.1 Algorithm Implementation

```

COMPUTEOPINIONRANK( $\mathcal{S}$ ,  $\mathcal{D}$ )
  ▷  $\mathcal{D}$  is the resource graph
  ▷  $\mathcal{S}$  is the social entity graph
1  $\mathcal{S} \leftarrow \text{COMPUTESOCIALRELEVANCE}(\mathcal{S})$ 
2 for each  $v \in \text{VERTEX}(\mathcal{D})$ 
   do
3    $v.source.relevance \leftarrow$ 
      $\text{NODES}(\mathcal{S})[v.source].relevance$ 
4  $\mathcal{D} \leftarrow \text{WEIGHTEDPAGERANK}(\mathcal{D})$ 

```

On the first step the social status of the authors of the information resource is calculated by using a social network measure and in particular the proximity prestige (2.5.1). Since well connected web pages tend to share the same amount of hyperlink targeting and coming back

from them (e.g through banner exchange) then their connectedness is also depicted in the social connection. For examples several persons are affiliated with a conference or an event so they provide a linking to that event but so does the event page as well denoting a social connection as well.

In the steps (2) and (3) of the algorithm execution the graph ( $D$ ) is signed with the results obtained by the *OWA* defuzization resulting a signed graph  $\acute{D} \equiv (D, E_A)$  where  $E_{AD} = OWA(u_D, u_{Dj})$  where  $j$  is the subsequent node that holds an adjacent connection with  $u$ . Having the resource graph signed with the defuzified valued of the opinion we then calculate the PageRank which is now *biased* by the social status that is attained by the author of the information resource.

### 5.3 Representing Fuziness

Representing the fuzzy variables and their defuzified values in our implementation was crucial. For that particular reason we used the FuzzyJ development kit supported by the National Research Council of Canada ([Orchard, 2004]). The API is written entirely in java and provides methods for accessing and defining fuzzy sets and the appropriate linguistic expressions (fuzzy variables).

Let us consider the directed graph  $\mathcal{S}$  of the social entities initialized in the step 0 of the algorithm. The average **in-degree** of the graph will be:

$$\langle k^{in} \rangle = \frac{1}{N} \bullet \sum_{i=1}^N k_i^{in} \quad (5.1)$$

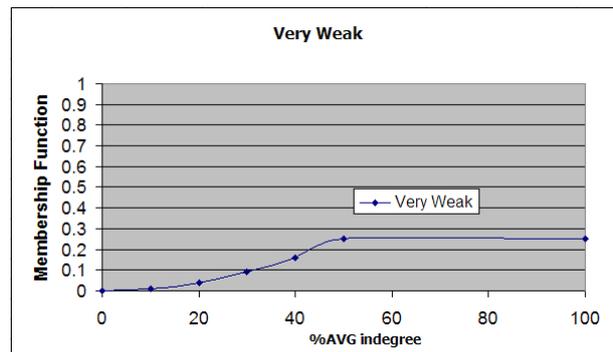
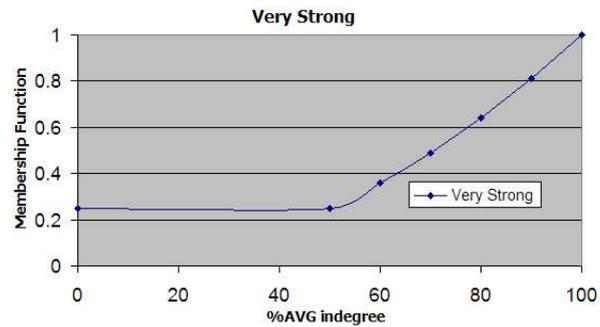
where  $k_i^{in}$  is the inner degree of the node  $i$  having  $i \in N$  We express the scale of the fuzzy values as a percentage of the average inner degree therefore we define a x-axis range of [0, 100] to our membership function coming as supplement to the already defined y-axis boundary of [0, 0.5, 1]

#### 5.3.1 Set and Variable Definition

We use the Fuzzyj toolkit to model the fuzzy set propositions in Java:

```
strength = new FuzzyVariable("strength", 0, 100, "%AVG(indegree)");
evidence = new FuzzyVariable("evidence", 0, 100, "%AVG(indegree)");

double xStrong[] = { 50, 100 };
double yStrong[] = { 0.5, 1 };
```



```
double xWeak[] = { 0, 50};
double yWeak[] = { 0, 0.5 };
```

Having defined the above range then the following linguistic variances of the variables can be obtained for the variable strength(the same values can represent the variances of evidence):

- Very Strong

```
FuzzyVariable      -> strength [ 0.0, 100.0 ] %AVG(indegree)
```

```
Linguistic Expression -> very strong
```

```
FuzzySet          -> { 0.25/50 0.36/60 0.49/70 0.64/80 0.81/90 1/100 }
```

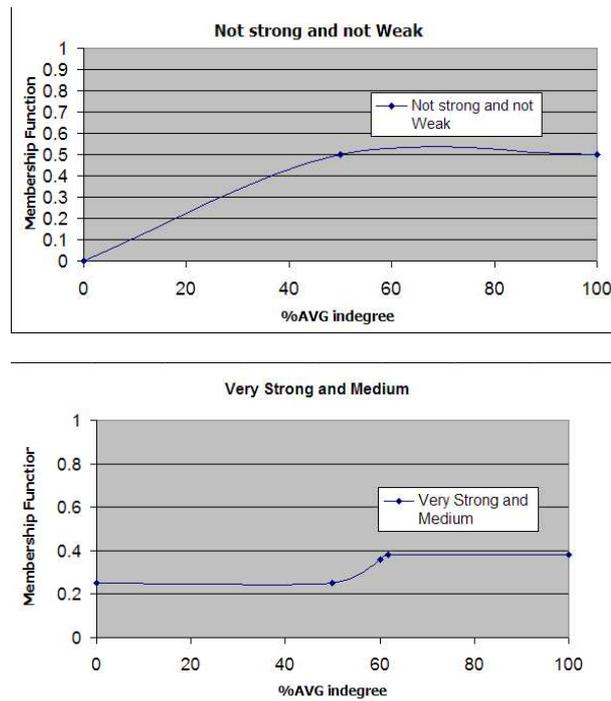
- Very Weak

```
FuzzyVariable      -> strength [ 0.0, 100.0 ] %AVG(indegree)
```

```
Linguistic Expression -> very weak
```

```
FuzzySet          -> { 0/0 0.01/10 0.04/20 0.09/30 0.16/40 0.25/50 }
```

- Not Strong and Not Weak



FuzzyVariable           -> strength [ 0.0, 100.0 ] %AVG(indegree)

Linguistic Expression -> not stron and not weak

FuzzySet               -> { 0.5/50 0/100 }

- Very Strong and Medium

FuzzyVariable           -> strength [ 0.0, 100.0 ] %AVG(indegree)

Linguistic Expression -> very strong and medium

FuzzySet               -> { 0.25/50 0.36/60 0.38/61.74 0/100 }

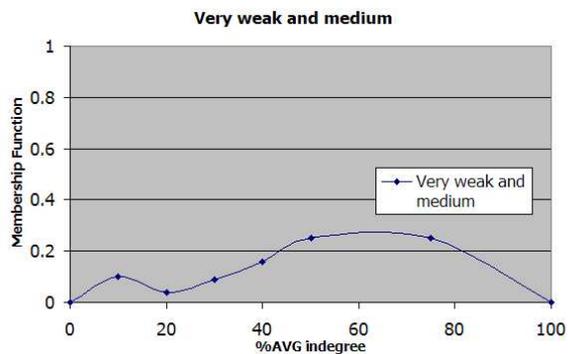
- Very weak and Medium

FuzzyVariable           -> strength [ 0.0, 100.0 ] %AVG(indegree)

Linguistic Expression -> very weak and medium

FuzzySet               -> { 0/0 0.01/10 0.04/20 0.09/30 0.16/40 0.25/50 0.25/75 0/100 }

The above plots provide an insight of the “possibilities” (degrees) that a node can be well connected, normally connected, less connected or not connected. The expressions of strength such as “very strong” give instance to the sociometric “stars” (Very Strong) and “isolates” (Very Weak) although different ranking may be resulted after the normalization process. To normalize the PageRank we used the ranker interface from the JUNG library.



## 5.4 Normalization

In order to sign the pairs in the graph we make a defuzization as described by the already defined operator (opinion). Then we use it to sign the edges on the document graph and recalculate the PageRank having as an input the social context which is provided by the operator.

```
Ranker PageRank = new PageRank(D,0.15,"opinion") ; PageRank.evaluate();
```

The dumping factor (0.15) remains the same for prior and posterior calculations in the graph  $\mathcal{D}$

In the next chapter we present a simulation based evaluation of the OpinionRank against the PageRank ranking algorithm using the proximity prestige of the authors as the distance measure (Author Rank). A subsequent discussion of the changes in rankings and the correlation between the two measures is provided.

## Chapter 6

# Validating OpinionRank

In that chapter we make a validation of the OpinionRank algorithm using simulation. For the construction of the network and the input data of the simulation model we used a generated random graph (in particular a Barabasi-Albert graph, see [Albert et al., 1999]) using the JUNG library. A presentation of the process and discussion of the results is presented in the sections that follow.

### 6.1 Why Simulation

Although simulation has been used in several fields only recently has found applications in social sciences. According to Axelrod ([Axelrod, 1998]) one of the primary reasons that simulation is not considered as a key tool in social sciences is the complexity which characterizes a social entity where present languages and simulation tools are lacking expressiveness. Furthermore such kind of complexity needs processing and storage capacity that will permit the effective execution of a model which in the case of contexts such as the World Wide Web demands enormous amounts of processing power and primary memory to evaluate and store the data obtained from the simulation model.

In our case we chose the paradigm of simulation mainly due to the complexity of implementing a web crawler and the hardware requirements for evaluating our algorithm in a large scale.

#### 6.1.1 Shortcomings

Although simulation provides an effective way of getting an insight of the model in the cases of information retrieval simulation is a tool that cannot be used to extract research results

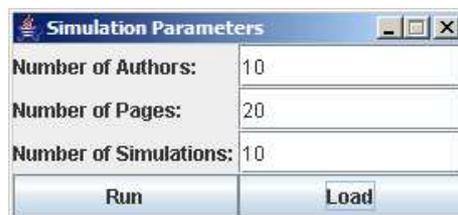
in a large scale due to the fact that precision as aforementioned, is a cognitive task of high complexity. This means that when a user submits a query in a search engine apart from the relevant pages that will be returned there are pages that satisfy the information need the most which are the pages that have highest precision. That kind of evaluation is difficult to be described by a stochastic or random simulation process therefore the case of getting experts to evaluate the results of information retrieval algorithms is considered the best case.

## 6.2 Model Construction

As aforementioned we used the Barabasi-Albert generator accessible from JUNG in order to obtain a small-world graph that can be used to analyze a hyperlink and social context. More specifically we used the implemented method `generateMixedRandomGraph` supplied with input parameters defined by the user for each rank in order to generate the graph and make the evaluation.

### 6.2.1 Input Parameters

The input parameters that are considered in our model :



Simulation Parameters	
Number of Authors:	10
Number of Pages:	20
Number of Simulations:	10
Run Load	

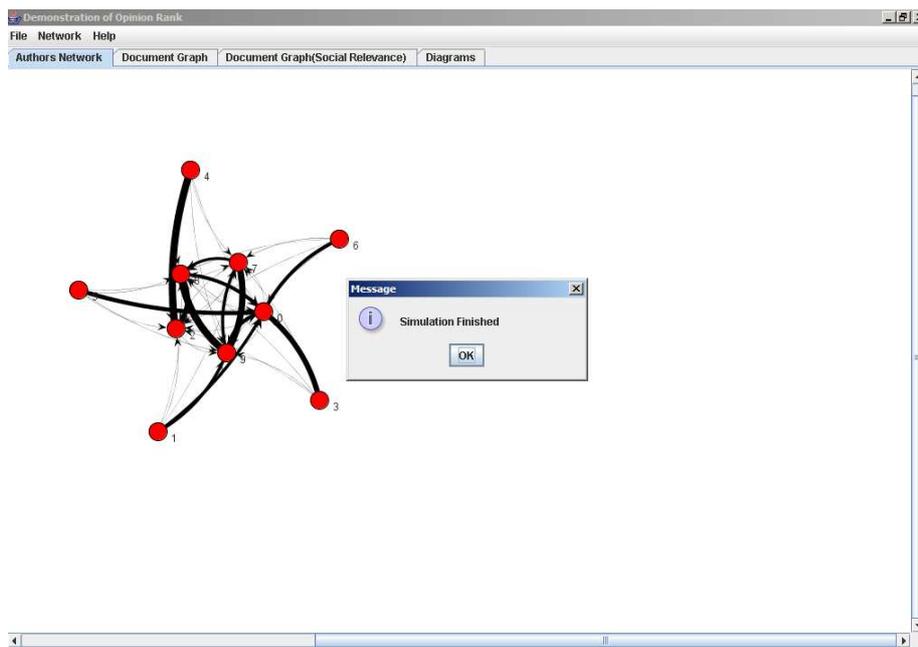
Figure 6.1: Entering simulation parameters

- The **Number of authors**: Which depicts the authorship network from which the social context of a resource is going to be evaluated
- The **Number of pages**: Which depicts the hyperlinked context of the information resource and
- The **Number of Simulations**: Which depicts the number of simulation rounds under which the model is going to run. To this is added an extra round which is not added to the model as to initiate the random number generator that we use for the construction of the pages networks.

The vertices obtained by the **Number of Pages** and **Number of Authors** are correlated as to have some authors to have ownership/affiliation with more than one pages.

### 6.2.2 Data Collection and Visualization

By using the developed environment we were able to obtain the results and have an abstract visualization of the rankings. We used the JUNG libraries to make an interactive visualization



**Figure 6.2:** Simulation Model

environment from which we navigate through the model and examine the results. In this demonstration application we have the following user interface functionalities:

**File > Parameters** Shows the relevant box for the input of model variables. By clicking the load button one can load the last known configuration values to the model.

**File > Run** Runs the model using the last known configuration values.

**Network > Write to EPS** Exports the current visualization to an encapsulated postscript image.

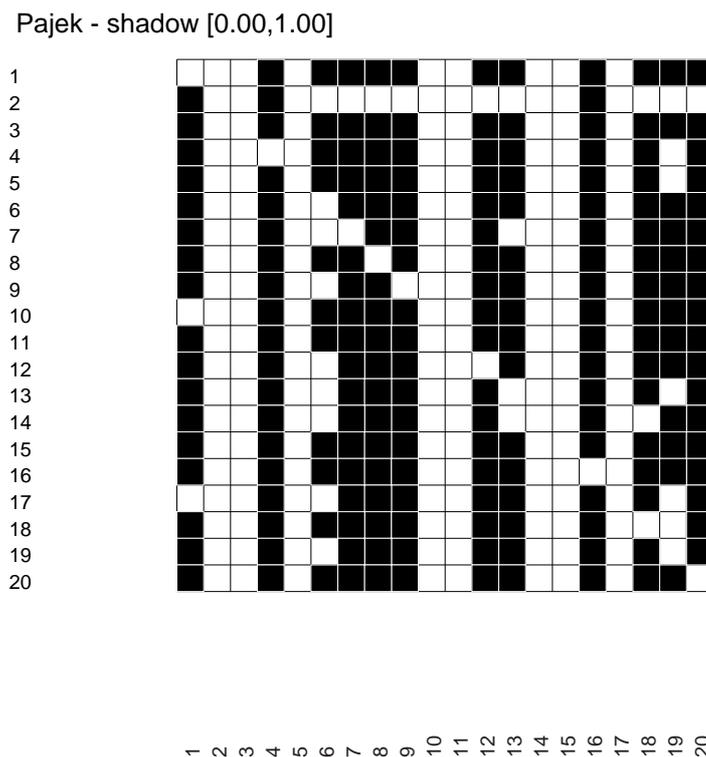
Apart from the Menu functionality after the simulation the program contains four different tabs which are:

**Authors Network** Contains a visualization of the social network of the authors. It provides a view of the link context

**Document Graph** Contains a visualization of the information resources graph as it is in the hyperlink context

**Document Graph (Social Relevance)** Contains a visualization of the *normalized* Document graph using the OpinionRank metric.

**Diagrams** Displays a chart where a comparison of the OpinionRank algorithm is displayed against the PageRank algorithm.



**Figure 6.3:** The sociomatrix of the resulted network after 10 rounds. The black boxes indicate whether this connection exists or not. Visualization was done using the Pajek software for network analysis ([Batagelj and Mrvar, 2002])

### 6.2.3 Running the model

After the parameters have been given, the simulation starts by generating two graphs as demanded by the algorithm.

- The Authors graph which is a random generated graph that represents the authors of the web page and their relational ties in a social network. Each author produces a number of web pages.
- The web pages graph is a random generated graph that denotes association between the web resources using hyperlinks.

A permutation is taking place in the authors graph and re-assigns randomly the relational ties and signs them with linguistic expressions of the fuzzy variables strength and evidence. Then the edges of the graph are iterated by a second process (normalization) where the Opinion (OWA) operator is combining the fuzzy values “strength” and “evidence”. Figure (6.2.2) shows the adjacency matrix (sociomatrix) of the patterns of the relational ties that are formed through the permutation on the initially generated random graph.

The tables 6.1 and 6.2 show the results after 10 rounds for a network of 10 authors and 20 pages.

After the simulation is done, a notification box appears in order to provide feedback that the simulation has finished (see figure 6.2). The screen now is filled with the resulted graphs that display the generated networks using a visualization layout. The layout that has been chosed is the Fruchterman-Raingold (FRLayout) which makes the arrangement in the plane as follows:

- The in-degree of each vertex is calculated along with the average in-degree of the graph
- A zero point is defined and a permutation arranges the nodes with the higher indegree in layers defined in the plane. The process is iterative as to result the layers with the highest indegree close to the zero point
- After the permutation is finished the vertices with the least indegree are arranged on the periphery while the vertices with the highest indegree are arranged in the core of the graph.

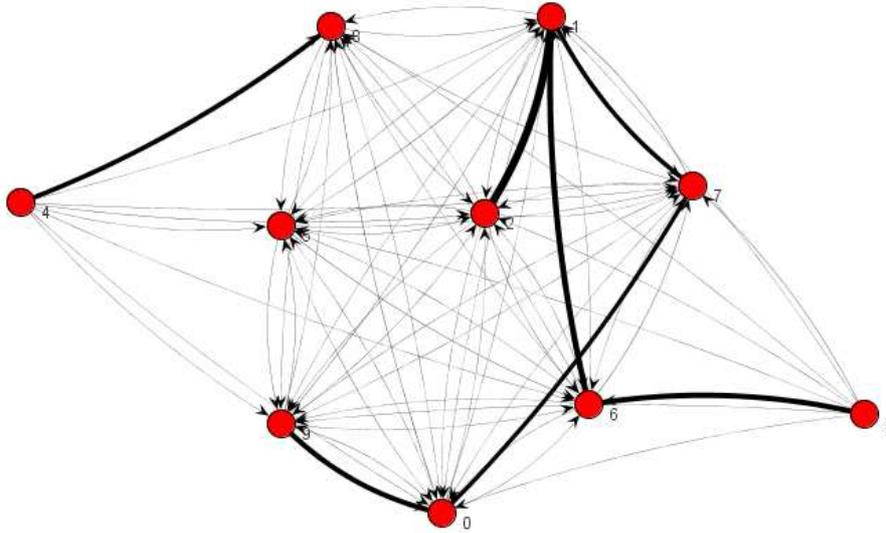
The edges are signed using the defuzzization value obtained by the Opinion operator. The edges then are decorated using a custom edge stroke from the JUNG interface **EdgeStrokeFunction**. For each edge the stroke is calculated as the result of the Opinion Operator multiplied with ten. For visualization reasons we chose to filter from the graph the strokes where the

<b>PageRank</b>		
Simulation Parameters	Authors: 10, Pages: 20 , Rounds: 10	
1	Vertex Id: 17	Rank 1: 0.091412
2	Vertex Id: 8	Rank 2: 0.084855
3	Vertex Id: 15	Rank 3: 0.084771
4	Vertex Id: 18	Rank 4: 0.077951
5	Vertex Id: 5	Rank 5: 0.077456
6	Vertex Id: 9	Rank 6: 0.077456
7	Vertex Id: 7	Rank 7: 0.077396
8	Vertex Id: 12	Rank 8: 0.075271
9	Vertex Id: 14	Rank 9: 0.075113
10	Vertex Id: 3	Rank 10: 0.075032
11	Vertex Id: 13	Rank 11: 0.065649
12	Vertex Id: 19	Rank 12: 0.057986
13	Vertex Id: 16	Rank 13: 0.027153
14	Vertex Id: 1	Rank 14: 0.007500
15	Vertex Id: 2	Rank 15: 0.007500
16	Vertex Id: 4	Rank 16: 0.007500
17	Vertex Id: 6	Rank 17: 0.007500
18	Vertex Id: 10	Rank 18: 0.007500
19	Vertex Id: 11	Rank 19: 0.007500
20	Vertex Id: 0	Rank 20: 0.007500
		Total: 1.000000

**Table 6.1:** Ranking of the node set obtained from PageRank

OpinionRank		
Simulation Parameters	Authors: 10, Pages: 20 , Rounds: 10	
1	Vertex Id: 15	Rank 1: 0.106737
2	Vertex Id: 9	Rank 2: 0.101469
3	Vertex Id: 17	Rank 3: 0.098875
4	Vertex Id: 14	Rank 4: 0.088891
5	Vertex Id: 12	Rank 5: 0.082498
6	Vertex Id: 18	Rank 6: 0.081753
7	Vertex Id: 3	Rank 7: 0.071187
8	Vertex Id: 8	Rank 8: 0.068435
9	Vertex Id: 13	Rank 9: 0.067778
10	Vertex Id: 5	Rank 10: 0.066218
11	Vertex Id: 7	Rank 11: 0.050258
12	Vertex Id: 19	Rank 12: 0.041627
13	Vertex Id: 16	Rank 13: 0.021772
14	Vertex Id: 1	Rank 14: 0.007500
15	Vertex Id: 2	Rank 15: 0.007500
16	Vertex Id: 4	Rank 16: 0.007500
17	Vertex Id: 6	Rank 17: 0.007500
18	Vertex Id: 10	Rank 18: 0.007500
19	Vertex Id: 11	Rank 19: 0.007500
20	Vertex Id: 0	Rank 20: 0.007500
		Total: 1.000000

**Table 6.2:** Ranking of the node set obtained from OpinionRank, notice the change of the ranking



**Figure 6.4:** Social network view of *OpinionRank*

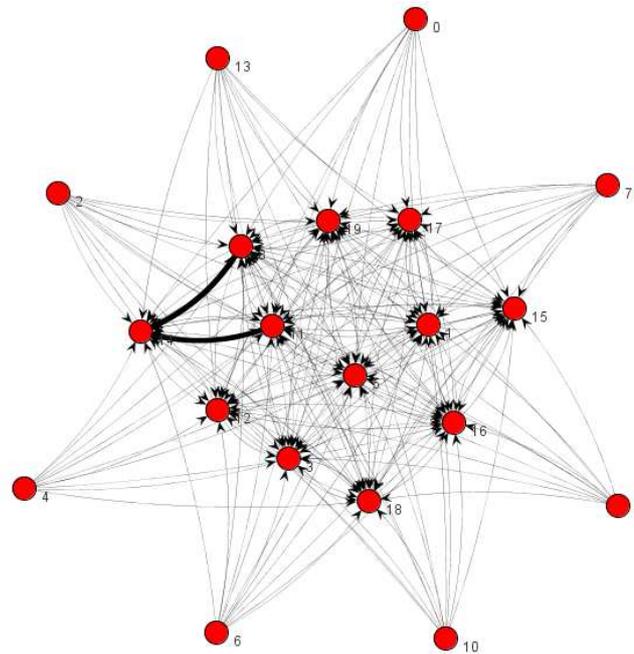
defuzified value is less than 0.4 which is the average value obtained in the network. Therefore we examine those that are above the average. In connection with our theoretical definitions in chapter 2 we attribute trustworthiness to those with the highest PageRank since they are recipients of the most positive choices in the network as a result obtained from the aggregation operator.

For the edges that are covered by the strokes we have that the thicker the stroke is then the stronger the opinion from the originated to the targeted vertex will be. For example in the figure 6.6 the opinion of the vertex “0” for the vertex “7” is less strong than the opinion of the vertex “2” for the vertex “1”.

### 6.3 Discussion

The simulation results are summarized in the tables 6.1, 6.2 and show the changes of the rankings with and without consideration of the Opinion in the graph structure.

In particular the figures 6.5 and 6.6 provide a visual exploration of the alternative rankings obtained by the PageRank and OpinionRank metrics subsequently. In the first case the results rank the nodes 17,8 and 15 as the most relevant however after the input of social context the ranking changes and ranks the nodes 15, 9, and 17 as the three most relevant. As can be seen in both cases node 17 is listed in the top which denotes that it’s hypertextual popularity depicts it’s social “prestige” as well.

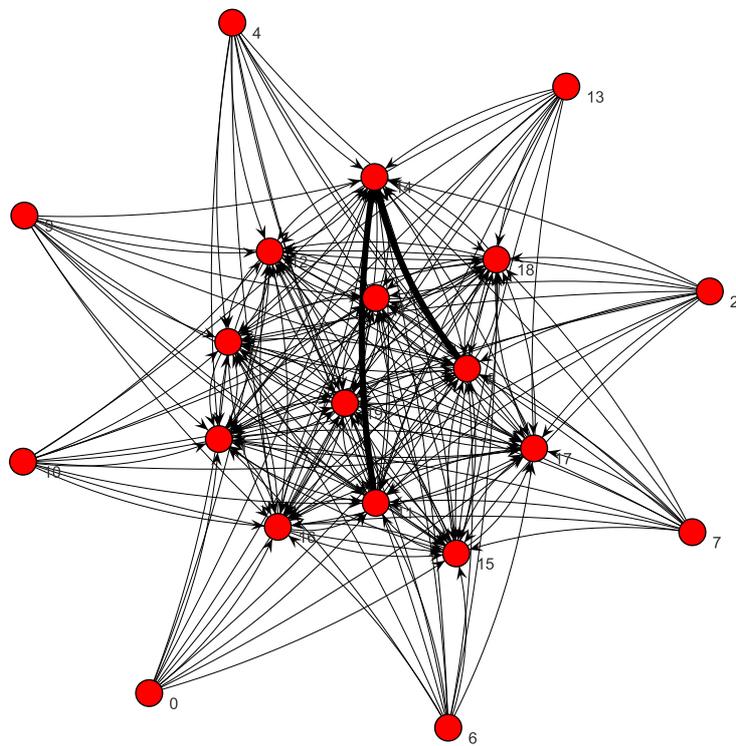


**Figure 6.5:** Opinion Evaluation depicted by the hyperlink context on the network of the web pages

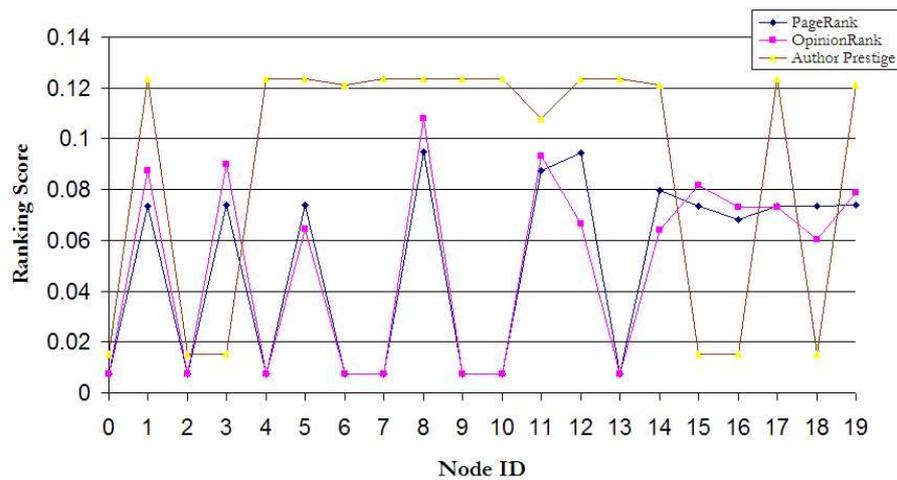
However as in the cases of nodes 15 and 9 the social prestige is not depicted in the hypertextual context which means that although the authors of this pages are considered as “prestigious” their works are not receiving much attention when it comes in the web.

The figure 6.7 shows a comparison between the two alternative rankings along with the variances of author prestige (AuthorRank) obtained from the social network and attributed to each web page that is ranked. The x-axis represents the nodes (node id number) and the y-axis the ranking score.

As can be seen in the first levels the social context attributes a variance to the original PageRank ranking for the first three ranking positions whereas the prestige drops but the hypertextual context (the popularity of the page) gives a high rank even where the social context is low. Furthermore in the ranking of the nodes the peaks on the social context show the variance of the membership in the fuzzy sets presented in chapter 6. In particular for the nodes 4-15 we clearly see a categorization to a level of strong relations that it reflects variances between the members (peaks in the PageRank and OpinionRank) as it is a directional property and not all the members of the set will have the reciprocity of their expressions.



**Figure 6.6:** Opinion Evaluation depicted by both the hypertextual and social context of the web pages and their affiliated authors



**Figure 6.7:** Evaluation Plot. The blue series represent the values obtained from the original PageRank ranking

## Chapter 7

# Conclusions and Future Research

### 7.1 General Conclusions

A model of Web page ranking combining link-based metrics with a basic account of social relations has been described. The model provides a tentative account for evidence and strength of relationships attributing them as the opinion of the individuals in the relational tie as well an external view of other entities in the social structure.

However relational ties and their instances are subject to a wide range of parameters and therefore the evaluation model could be more complex. Furthermore this thesis considered only a small part of those instances expressed in social network theory as well as their representation on a metric that could express them in a single value thus attribute it to the hyperlink context.

As regarding our formulated research questions we have come across some important remarks:

- Imprecise modeling of the relational ties can be done by aggregations of the properties that characterize them. In our case we expressed the relational tie using an aggregation operator (see definition on Chapter 5).
- Web Page relevance is biased by the social context. However **the relation is not reciprocal** which means that even if the author is among the most prestigious entities in the social context this doesn't necessarily imply that his/her productions are the most popular as well.
- Such kind of a metric demands input from the social context of the affiliated authorship obtained through the hypermedia context. Nonetheless is not always clear if the relational tie is between two or more entities. For instance in our simulation experiment we

attributed the authorship only to one person whereas in the web there are pages that often depict the opinion of more than one author. From the other hand people tend to author pages that are sensitive to the social context such as the case of blogs. We believe that our metric could ideally work as a ranking model for blogs and other single entity instances of web presence

Another concluding remark is that the defuzzification of the imprecise expressions is something that may need to be evaluated further as to approach several variances of social contexts whereas the obtained ranking is not the same. However in our case the output provided a clear bias in the rankings depicting the most relevant in both contexts. Imprecise expression of the variances of strength and evidence over the directional relations expressed in the network provide us a different viewpoint on the evaluation of the opinion as a metric of social and hyperlink prominence. Depending on the semantics of strength and evidence an opinion can have different implications for ranking as in the case of affective ties where for instance the algorithmic proposition should be different as to penalize the negative ties in favor of the positive (structural balance).

## 7.2 Lessons Learned

Due to the writing and the related work carried out on the purpose of this thesis there were several encounters which we believe it is worth noting:

- The original idea was to create a publication platform where each member could construct and express its social profile through FOAF metadata. However this added development complexity to the project due to the fact that FOAF vocabulary needed to be extended to capture the extra properties required for the quantification of the relationship therefore the direction of using simulation as an alternative time effective approach for validation was adopted.
- The incorporation of fuzzy logic came after long thoughts since the topic was completely unknown to the author. Nonetheless the concept of fuzzy logic and soft-computing in general is something that has transformed today's technological appliances (In Japan the fuzzy logic industry has an annual revenue of 4 billion . Social applications of fuzzy logic although available in the literature (see [Ragin, 2000]) were difficult to find and consult since they cover a variety of topics without much diversification.

## 7.3 Future Research

Since this thesis was grounded on three different fields namely Sociology, Information Retrieval and Fuzzy Logic we have recognized some future research that needs to be carried to the extend of both three fields.

**Role of Structural Balance** Current models of Structural Balance work on clusters resulted from the main sociogram. Fore example the hubness and the authoritativeness of a particular node should be considered as an indicator of common interest (eg. web communities about aerospace should provide a fair amount of links to authorities such as Airbus or Boeing). Following the approach initiated by Gibson ([Gibson et al., 1998]) it would be interesting to evaluate ranking in small communities such as the Slashdot.org or Orkut.com.

**Fuzzy Expressions of Centrality** Prominence is not something that is always clear since a real world social structure is far more complex than the structures deployed in this thesis.

**Evaluation of Transitivity on FOAF Metadata** The FOAF vocabulary provides a way of expressing transitive properties (I'm a friend-of-a-friend) between interconnected entities. It would be interesting to explore the extend to which this transitive property remains.

Nonetheless evaluation and extension of this thesis is also subject to data that have to be gathered from the web in order to come with a concrete proposition. Currently we are targeting to submit this thesis to a conference in order to get feedback from the research community members.



## Appendix A

# Running the Simulation Environment

In order to run the simulation environment you need to have the JAVA runtime environment in your computer (versions 1.4.x and greater). The zip file that contains the simulator can be downloaded from:

<http://www.dsv.su.se/~x04-nko/thesis/opinionrank-simulator.zip>

Once you download the file unzip it to a folder and double click the run.bat (see figure A.1)

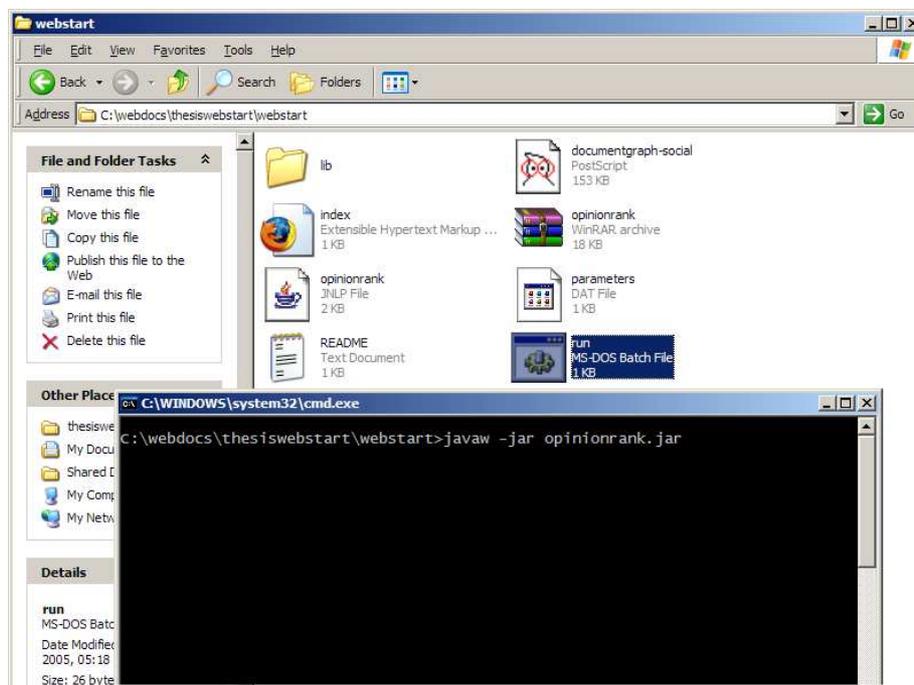


Figure A.1: Launching the simulation environment



## Appendix B

# Index of Abbreviations

**Actor:** The term “actor” refers to a person or a group that is subject of a social relation. The actor is represented by a vertex in a social network

**Clique:** A clique is the maximal complete subnetwork containing three or more vertices.

**Sociogram:** The sociogram is the planar layout representation of a graph that represents a social network.

**Graph:** A graph is set of vertices and a set of lines between pairs of vertices.

**Directed Graph:** A directed graph is a graph where the vertices are connected with directional lines (arcs).

**Undirected Graph:** An undirected graph is a graph that the lines have no direction, thus the connection they represent is reciprocal.

**Aggregation:** Aggregation is the property of a system to summarize the different parameters that are involved during the operation, into a representative value.

**Eigenvector:** The eigenvector is a non-zero vector that when is operated by an operator results a scalar multiple of itself.

**FOAF:** The Friend-of-a-Friend vocabulary is a metadata schema for describing social relations based on RDF

**Bivalence:** Bivalence is a principle in logic that asserts that a proposition will be either true or false.

**Precision:** In information retrieval “precision” is a metric that describes the accuracy of the information retrieval mechanism in terms of the percentage of the actually relevant documents than those judged relevant by the mechanism. Precision is usually an outcome of cognitive evaluation of the results retrieved.

## Appendix C

# Literature Collected

The literature collected and studied through the preparation of this thesis is available on CiteUlike which is a free service for collecting and managing academic papers. The respective url under which the references can be found is: <http://www.citeulike.org/user/nkorf/tag/thesis>



# Bibliography

- [Albert et al., 1999] Albert, R., Jeong, H., and Barabasi, A.-L. (1999). The diameter of the world wide web. *Nature*, 401:130–131.
- [Aronson et al., 2004] Aronson, E., Wilson, T. D., and Akert, R. M. (2004). *Social Psychology*. Prentice Hall, fifth edition.
- [Axelrod, 1998] Axelrod, R. (1998). Advancing the art of simulation in the social sciences. *Complexity*, 3(2):16–22.
- [Batagelj and Mrvar, 2002] Batagelj, V. and Mrvar, A. (2002). *Pajek - Analysis and Visualization of Large Networks*, volume 2265, pages 477+. Springer Verlag.
- [Bavelas, 1948] Bavelas, A. (1948). A mathematical model of group structure. *Human Organizations*, 7:16–30.
- [Berger and Luckmann, 1967] Berger, P. L. and Luckmann, T. (1967). *The Social Construction of Reality : A Treatise in the Sociology of Knowledge*. Anchor.
- [Berkowitz, 1988] Berkowitz, S. D. (1988). Markets and market areas: Some preliminary formulations. In Wellman, B. and Berkowitz, S. D., editors, *Social Structures: A network approach*, pages 261–303. Cambridge University Press, Cambridge, England.
- [Berners-Lee et al., 2001] Berners-Lee, T., Hendler, J., and Lassila, O. (2001). The semantic web. *The Scientific American*.
- [Bollobas, 2001] Bollobas, B. (2001). *Random Graphs*. Cambridge University Press.
- [Brickley and Miller, 2005] Brickley, D. and Miller, L. (2005). FOAF vocabulary specification.
- [Brin and Page, 1998] Brin, S. and Page, L. (1998). The anatomy of a large-scale hypertextual web search engine. In *Proceedings of the Seventh International Conference on World Wide Web*.  
Web.

- [Bron and Kerbosch, 1973] Bron, C. and Kerbosch, J. (1973). Algorithm 457: finding all cliques of an undirected graph. *Communications of the ACM*, 16(9):575–577.
- [Brouwers, 2005] Brouwers, L. (2005). Micropox: A large scale and spatially explicit micro-simulation model for smallpox transmission. In *Proceedings of the International Conference on Health Sciences Simulation*, New Orleans.
- [Brown and Duguid, 2002] Brown, J. S. and Duguid, P. (2002). *The Social Life of Information*. Harvard Business School Press.
- [Burt, 1982] Burt, R. S. (1982). *Toward a structural theory of action : network models of social Structure, Perception, and Action*. Academic Press, New York.
- [Cartwright and Harary, 1956] Cartwright, D. and Harary, F. (1956). Structural balance: A generalization of heiders theory. *Psychological Review*, 63:277–292.
- [Cartwright and Harary, 1970] Cartwright, D. and Harary, F. (1970). Ambivalence and indifference in generalizations of structural balance. *Behavioral Science*, 15:497–513.
- [Cho and Garcia-Molina, 2000] Cho, J. and Garcia-Molina, H. (2000). The evolution of the web and implications for an incremental crawler. In *Proceedings of the Twenty-sixth International Conference on Very Large Databases*.
- [Colamen et al., 1957] Colamen, J. S., Katz, E., and Menzel, H. (1957). The diffusion of an innovation among physicians. *Sociometry*, 20:253–270.
- [Daley and Gani, 2001] Daley, D. J. and Gani, J. (2001). *Epidemic Modelling : An Introduction (Cambridge Studies in Mathematical Biology)*. Cambridge University Press.
- [DeRose et al., 2001] DeRose, S., Maler, E., and Orchard, D. (2001). Xml linking language (xlink) version 1.0. *W3C Recommendation*.
- [Doreian and Fararo, 1985] Doreian, P. and Fararo, T. J. (1985). Structural equivalence in a journal network. *Journal of the American Society for Information Science and Technology*, 36:28–37.
- [Erdos and Renyi, 1960] Erdos, P. and Renyi, A. (1960). On the evolution of random graphs. In *Publications of the mathematical institute of the Hungarian Academy of Sciences*, volume 5, pages 17–61.

- [Faloutsos, 1985] Faloutsos, C. (1985). Access methods for text. *ACM Computing Surveys*, 17(1):49–74.
- [Faloutsos et al., 1999] Faloutsos, M., Faloutsos, P., and Faloutsos, C. (1999). On power-law relationships of the internet topology. In *SIGCOMM '99: Proceedings of the conference on Applications, technologies, architectures, and protocols for computer communication*, volume 29, pages 251–262, New York, NY, USA. ACM Press.
- [Frakes and Baeza-Yates, 1992] Frakes, W. B. and Baeza-Yates, R. (1992). *Information Retrieval: Data Structures and Algorithms*. Prentice Hall PTR.
- [Freeman, 2000] Freeman, L. (2000). Social network analysis: Definition and history. In Kazdan, A. E., editor, *Encyclopedia of Psychology*, volume 6, pages 350–351. Oxford University Press, New York.
- [Friedkin, 1998] Friedkin, N. E. (1998). *A Structural Theory of Social Influence (Structural Analysis in the Social Sciences)*. Cambridge University Press.
- [Frienklin, 1984] Frienklin, N. E. (1984). Structural cohesion and equivalence explanations of social homogeneity. *Sociological Methods and Research*, 12:235–261.
- [Gibson et al., 1998] Gibson, D., Kleinberg, J., and Raghavan, P. (1998). Inferring web communities from link topology. proc. 9th acm conference on hypertext and hypermedia, 1998. In *Proceedings of the 9th ACM Conference on Hypertext and Hypermedia*. ACM.
- [Granovetter, 1982] Granovetter, M. (1982). The strength of weak ties: A network theory revisited. In Marsden, P. V. and Lin, N., editors, *Social Structure and Network Analysis*. Sage Publications, Beverly Hills, CA.
- [Gyongyi et al., 2004] Gyongyi, Z., Garcia-Molina, H., and Pedersen, J. (2004). Combating web spam with trustrank. In *Proceedings of the 30th International Conference on Very Large Data Bases (VLDB)*.
- [Harary, 1969] Harary, F. (1969). *Graph Theory*. Perseus Books.
- [Heider, 1946] Heider, F. (1946). Attitudes and cognitive organization. *Journal of Psychology*, 21:107–112.
- [Katz, 1953] Katz, L. (1953). A new status index derived from sociometric analysis. *Psychometrika*, 18:39–43.

- [Kelley, 1971] Kelley, H. H. (1971). Attribution in social interaction. In Jones, E. E., Kanouse, D. W., Kelley, H. H., Nisbett, R. E., Valins, S., and Weiner, B., editors, *Attribution: Perceiving the causes of behavior*, pages 1–21. General Learning Press, Morristown, NJ.
- [Kleinberg, 1999] Kleinberg, J. M. (1999). Hubs, authorities, and communities. *ACM Comput. Surv.*, 31(4es).
- [Klyne and Carroll, 2002] Klyne, G. and Carroll, J. J. (2002). Resource description framework (RDF): Concepts and abstract syntax. *W3C Recommendation*.
- [Kohler, 1947] Kohler, W. (1947). *Gestalt Psychology*. Liveright, New York.
- [Krackhardt, 1992] Krackhardt, D. (1992). The strength of strong ties: The importance of philos in organizations. pages 216–239.
- [Kruskal and Wish, 1978] Kruskal, J. B. and Wish, M. (1978). *Multidimensional Scaling (Quantitative Applications in the Social Sciences)*. SAGE Publications.
- [Kumar et al., 2000] Kumar, R., Maghoul, F., Raghavan, P., Rajagopalan, S., Stata, R., Tomkins, A., and Wiener, J. L. (2000). Graph structure in the web. *Computer Networks*, 33(1-6):309–320.
- [Laumann et al., 1983] Laumann, E. O., Marsden, P. V., and Prensky, D. (1983). The boundary specification problem in network analysis. In Burt, R. S. and Minor, M. J., editors, *Applied Network Analysis*, pages 18–34. Sage Publications, London.
- [Leavitt, 1951] Leavitt, H. J. (1951). Some effects of certain communication patterns on group performance. *Journal of Abnormal and Social Psychology*, 46(1):38–50.
- [Leifer and White, 1987] Leifer, E. M. and White, H. C. (1987). A structural approach to markets. In Mizruchi, M. S. and Schwarz, M., editors, *Intercorporate Relations: The Structural analysis of business*, pages 85–108. Cambridge University Press, Cambridge, UK.
- [Leuf and Cunningham, 2001] Leuf, B. and Cunningham, W. (2001). *The Wiki Way: Collaboration and Sharing on the Internet*. Addison-Wesley Professional.
- [Leydesdorff, 2001] Leydesdorff, L. (2001). *The Challenge of Scientometrics: The Development, Measurement, and Self-Organization of Scientific Communications*. Universal Publishers.

- [Licklider et al., 1968] Licklider, J., Taylor, R., and Herbert, E. (1968). The computer as a communication device. *Science and Technology: For the Technical Men in Management*, pages 21–31.
- [Lotka, 1926] Lotka, A. J. (1926). The frequency distribution of scientific productivity. *Journal of the Washington Academy of Sciences*, 16:317–323.
- [March and Lave, 1975] March, J. G. and Lave, C. A. (1975). *Introduction to Models in the Social Sciences*. Harpercollins College Div.
- [Mathes, 2004] Mathes, A. (2004). *Folksonomies: Cooperative classification and communication through shared metadata*. PhD thesis, Library and Information Science.
- [Milgram, 1967] Milgram, S. (1967). The small world problem. *Psychology Today*, 61.
- [Moreno, 1946] Moreno, J. L. (1946). Sociogram and sociomatrix. *Sociometry*, 9:348–349.
- [Moreno and Jennings, 1945] Moreno, J. L. and Jennings, H. H. (1945). Sociometric measurements of social configurations based on deviations from chance. NY.
- [Mullins, 1973] Mullins, N. C. (1973). *Theories and Theory Groups in Contemporary American Sociology*. Harper and Row, New York.
- [Newman, 2003] Newman, M. E. J. (2003). The structure and function of complex networks. *SIAM Review*, 45(2):167–256.
- [Orchard, 2004] Orchard, B. (2004). Fuzzyj toolkit and fuzzyjess. Technical report, National Research Council of Canada (NRC).
- [Page et al., 1998] Page, L., Brin, S., Motwani, R., and Winograd, T. (1998). The Pagerank Citation Ranking: Bringing Order to the Web. Technical report, Stanford Digital Library Technologies Project.
- [Pattison, 1993] Pattison, P. (1993). *Algebraic Models for Social Networks (Structural Analysis in the Social Sciences)*. Cambridge University Press.
- [Pinski and Narin, 1976] Pinski, G. and Narin, F. (1976). Citation influence for journal aggregates of scientific publications: Theory, with application to the literature of physics. *Information Processing and Management*, 12:297–312.

- [Prete et al., 2003] Prete, C. D., McArthur, J. T., Villars, R. L., , and Reinsel, D. (2003). Industry developments and models, disruptive innovation in enterprise computing: storage.
- [Ragin, 2000] Ragin, C. C. (2000). *Fuzzy-Set Social Science*. University Of Chicago Press.
- [Raub and Weesie, 1990] Raub, W. and Weesie, J. (1990). Reputation and efficiency in social interactions: An example of network effects,. *American Journal of Sociology*, 96:626–654.
- [Rogers, 1979] Rogers, E. M. (1979). Network analysis of the diffusion of innovations. In Holland, P. W. and Leinhardt, S., editors, *Perspectives on Social Network Research*, pages 137–164. Academic Press, New York.
- [Romney and Faust, 1982] Romney, A. K. and Faust, K. (1982). Predicting the structure of a communication network from recalled data. *Social Networks*, 4:285–304.
- [Ross et al., 1977] Ross, L., Amabile, T. M., and Steinmetz, J. L. (1977). Social roles, social control, and biases in social-perception processes. *Journal of Personality and Social Psychology*, 35:485–494.
- [Sabidussi, 1966] Sabidussi, G. (1966). The centrality index of a graph. *Psychometrika*, 31:581–603.
- [Sola Pool and Kochen, 1971] Sola Pool, I. and Kochen, M. (1971). Contacts and influence. *Social Networks*, 1(1):5–51.
- [Wasserman et al., 1994] Wasserman, S., Faust, K., and Iacobucci, D. (1994). *Social Network Analysis : Methods and Applications (Structural Analysis in the Social Sciences)*. Cambridge University Press.
- [Yager, 1988] Yager, R. R. (1988). On ordered weighted averaging aggregation operators in multicriteria decisionmaking. *IEEE Transactions on Systems, Man and Cybernetics*, 18(1):183–190.
- [Zadeh, 1969] Zadeh, L. A. (1969). Fuzzy sets. *Information and Control*, 8:338–353.