



RESEARCH ARTICLE

On Some Aspects of Link Analysis and Informal Network in Social Network Platform

Subrata Paul

Department of Computer Science and Engineering, M.I.T.S. Rayagada, Odisha 765017, INDIA

subratapaulcse@gmail.com

Abstract— This paper presents a review on the two important aspects of Social Network, namely Link Analysis and Informal Network. Both of these characteristic plays a vital role in analysis of direction of information flow and reliability of the information passed or received. They can be easily visualized or can be studied using the concepts of graph theory. We have deliberately omitted discussing about general definitions of social network and representing the relationship between actors as a graph with nodes and edges. This paper starts with a formal definition of Informal Network and continues with some of its major aspects. Moreover its application is explained with a real life example of a college scenario. In the later part of the paper, some aspects of Informal Network have been presented. The similar kind of real life scenario is also being drawn out here to represent application of informal network. Lastly, this paper ends with a general conclusion on these two topics.

Key Terms: - Social Network; Link Analysis; Information Overload; Formal Network; Informal Network

I. INTRODUCTION

Link analysis is a data-analysis technique used to evaluate relationships (connections) between nodes. Relationships may be identified among various types of nodes (objects), including organizations, people and transactions. Link analysis has been used for investigation of criminal activity (fraud detection, counterterrorism, and intelligence), computer security analysis, search engine optimization, market research and medical research. Link analysis is one of many factors considered by web search engines in computing a composite score for a web page on any given query. Link analysis is in many ways similar to social network analysis (SNA)—both talk about relationships in terms of nodes and edges and both try to derive the idea of who is more important in a network by analyzing the whole network, not individual events.

Informal networks by is the part of social network, where there exist no defined or formal structure between nodes (people belonging to the network). It's more like creating an ad-hoc relationship. People responding to new work directives may act instantly to update their procedures, or may take days to organize themselves in line with company policy. Informal network has its own importance. Informal networks by definition have no structure. People responding to new work directives may act instantly to update their procedures, or may take days to organize themselves in line with company policy.

II. COMPONENTS OF LINK ANALYSIS

Link analysis can be broken down into two components — link generation and utilization of the resulting linkage graph.

A. *Link Generation*

Link generation is the process of computing the links, link attributes and node attributes. There are several different ways to define links. The different approaches yield very different linkage graphs. A key aspect in defining a link analysis is deciding which representation to use.

B. *Explicit Links*

A link may be created between the nodes corresponding to each pair of entities in a transaction. For example, with a call detail record, a link is created between the originating telephone number and the destination telephone number. This is referred to as an explicit link.

C. *Aggregate Links*

A single link may be created from multiple transactions. For example, a single link could represent all telephone calls between two parties, and a link attribute might be the number of calls represented. Thus, several explicit links may be collapsed into a single aggregate link.

D. *Inferred Relationships*

Links may also be created between pairs of nodes based on inferred strengths of relationships between them. These are sometimes referred to as soft links, association links, or co-occurrence links. Classes of algorithms for these computations include association rules, Bayesian belief networks and context vectors. For example, a link may be created between any pair of nodes whose context vectors lie within a certain radius of one another. Typically, one attribute of such a link is the strength of the relationship it represents. Time is a key feature that offers an opportunity to uncover linkages that might be missed by more typical data analysis approaches. For example, suppose a temporal analysis of wire transfer records indicates that a transfer from account A to person X at one bank is temporally proximate to a transfer from account B to person Y at another bank. This yields an inferred link between accounts A and B. If other aspects of the accounts or transactions are also suspicious, they may be flagged for additional scrutiny for possible money laundering activity.

A specific instance of inferred relationships is identifying two nodes that may actually correspond to the same physical entity, such as a person or an account. Link analysis includes mechanisms for collapsing these to a single node. Typically, the analyst creates rules or selects parameters specifying in which instances to merge nodes in this fashion [28] [23] [6].

E. *Utilization*

Once a linkage graph, including the link and node attributes, has been defined, it can be browsed, searched or used to create variables as inputs to a decision system.

III. ISSUES WITH LINK ANALYSIS

Link analysis often makes information accessible that is not apparent from any single data record, which is an advantage. However, it has its own weakness. Link analysis is as much an art as a science, and just configuring a link analysis can be a major endeavour. One of the major issues of link analysis is information overload. This issue is discussed in the next sections.

Information Overload

With the vast amounts of data and information that are stored electronically, users are confronted with multiple unrelated sources of information available for analysis. Data analysis techniques are required to make effective and efficient use of the data. We studied on the classification made by Palshikar. He had classified data analysis techniques into two categories – statistical (models, time-series analysis, clustering and classification, matching algorithms to detect anomalies) and artificial intelligence (AI) techniques (data mining, expert systems, pattern recognition, machine learning techniques, neural networks).

Bolton & Hand define statistical data analysis as either supervised or unsupervised methods. Supervised learning methods require that rules are defined within the system to establish what is expected or unexpected behavior. Unsupervised learning methods review data in comparison to the norm and detect statistical outliers. Supervised learning methods are limited in the scenarios that can be handled as this method requires that training rules are established based on previous patterns. Unsupervised learning methods can provide detection of broader issues, however, may result in a higher false-positive ratio if the behavioral norm is not well established or understood.

Data itself has inherent issues including integrity (or lack of) and continuous changes. Data may contain “errors of omission and commission because of faulty collection or handling, and when entities are actively attempting to deceive and/or conceal their actions highlights incompleteness (inevitability of missing data or links), fuzzy boundaries (subjectivity in deciding what to include) and dynamic changes (recognition that data is ever-changing) as the three primary problems with data analysis.

Once data is transformed into a usable format, open texture and cross referencing issues may arise. Open texture was defined by Waismann as the unavoidable uncertainty in meaning when empirical terms are used in different contexts. Uncertainty in meaning of terms presents problems when attempting to search and cross reference data from multiple sources [18].

IV. A REAL LIFE EXAMPLE OF LINK ANALYSIS

We have already discussed about link analysis and literature related to it in earlier section. Link analysis is a data-analysis technique used to evaluate relationships (connections) between nodes. Relationships may be identified among various types of nodes (objects), including organizations, people and transactions. Link analysis has been used for investigation of criminal activity (fraud detection, counterterrorism, and intelligence), computer security analysis, search engine optimization, market research and medical research. Let us assume the case of an educational institution where 4 levels of entities are described as Director in the first, HOD’s in the second, faculties in the third and students in the last level. Let us also assume the case where a particular instruction such as the deadline of project submission of the final year M.Tech students and some general instructions regarding the project are to be conveyed to the students from the highest level. The Director does not have a direct link with the students. So the instruction is to be conveyed step by step from one level to another so that it finally reaches the students.

The director has a direct link with the HOD’s of all the departments of the institution and so he can easily convey the set of instructions to the HOD’s of all departments. This forms the first level of relationship where the information is passed from one level to another. In the next level, the HOD has a direct link with faculties of his concerned department under him. So now he can easily convey the information that he has received from the Director to his subordinate faculties. During the process of passing the information any new additions or deletions to the original set of instructions can be made. The faculty has a direct link with the students and so he conveys the instruction to the students in order to guide them for submitting the project at an appropriate date. This forms the final level of the relationship and there lies the possibility that the instructions can be added, deleted or modified before it finally reaches the students. Thus we visualize that even though the Director has no direct link with the students, the set of instructions generated by him is conveyed to the students through an orderly manner in a stepwise process.

Let us consider a different case where a particular student feels that a specific instruction with regards to project submission is too hard to follow and needs to be immediately modified. Except for some exceptional instances, the student can’t go directly to the Director to speak regarding his concerned problem. He has to inform the faculty since he has a direct link with him only. The faculty may convey the matter to the HOD, if he feels so or may resolve the matter by him also. The similar situation may arise with the HOD also.

Thus, in our example for Link Analysis, we had observed about the connections between highest entities (i.e. director) to the lowest entity (i.e. students). It allows us to study on direction of flow of information and if the information is modified, at which link. Even, the same can be observed for the information flow in reverse direction.

V. THE CONTENTS OF INFORMAL NETWORK

All The types of links between nodes in a social network can be described and analysed in a variety of ways. In the following section, a synthesis and categorization of the different approaches will be presented: This categorization will be based somewhat on the division proposed by Tichy & Tushman (1979) [27]: (1) The transactional content, (2) the nature of the links, and (3) the structural characteristics.

The transactional content

The connection between two nodes can have very different contents and implications. By having different focal points, it is possible to identify several alternative but overlapping approaches to defining the content typologies. Again it is paramount to specify that the same two actors in a social network can have several different transactional contents at different times and simultaneously. Monge & Eisenberg (1987) propose a

grouping of the contents in social networks into the following self-explaining typology [22]: *Expression of affect, influence attempt, exchange of information, and exchange of goods and services.*

Farace et al. (in Monge & Eisenberg, 1987) have chosen a somewhat different approach in that they identified the following three types of messages: *Production*-related messages stem from the need to get things done in a work context. *Innovation* messages are more proactive and are centred on solving organizational problems and improving the ways things are done. *Maintenance* messages revolve around the socio psychological needs of the individual as described previously [22].

A taxonomy directly linking to the division between formal and informal networks together with the corporate culture is proposed by Tichy & Tushman (1979) [27]. The *Technical* approach deals with work-related issues; the *Political* approach relates to individual and group goals, whereas the *Cultural* approach reaches into the implicit, tacit and deeper meanings and shared values in the organization. For more information on the interactions between these three levels see Culture and Informal Networks.

The advice network can be used to determine who has the technical or professional power in an organization. *The trust relations* reveal ties of friendship and affection, whereas *the communication network* is a strong indicator of the overall information flow in the organization.

Finally Torenvlied (1998) has proposed the following divisions, based on Ibarra (1992): *Instrumental* relations used in daily interactions based on short-term goals, *authority* relations based on power and influence and finally *intimate* relations [11].

As stated in the beginning of this section, those divisions are overlapping and – not surprisingly – rather alike. Table 1 is an attempt to classify most of the above terms into a unifying division.

TABLE I
Four main contents of Informal networks

Affect Friendships, trust and intimate relations	Political Influence, power, authority
Production Advice, exchange of technical/instrumental knowledge and innovation	Cultural Communication and flow of information

VI. PROPERTIES OF THE LINKS

In order to take into account the different ways in which individuals enter relations, a measure of *symmetricality* is introduced. In a situation where two colleagues chat together during lunch, there is a symmetrical relation, whereas the supervisor-subordinate relationship is inherently asymmetrical.

This leads to the issue of power in a social relation and organizational politics, which is a matter deserving a whole section, and will not be covered as such in this paper.

Related to that is the measure of *reciprocity* which determines not the definition of the link but the degree to which individuals report the same intensities with each other for a content area (Tichy & Tushman, 1979) [27] . Unrelated to the degree of symmetry, a link between nodes can be reciprocal or not. If both supervisor and subordinate report that the supervisor “gives instruction to” the subordinate once or twice a week, the relation is reciprocal but not symmetrical.

The term *multiplexity* refers to the extent to which each individual has different roles and different networks within the organization. Those different networks are closely related to the contents described in the previous section, and it is important to keep in mind that the same two individuals can easily have more than one relation to each other. When describing and dealing with social networks, this is a fact that is often a major source of difficulty.

Clarity of expectations is the general measure of whether or not the two parties in a relation agree on the appropriate behaviour between them. Seen together with the multiplexity of many relations, it follows that the appropriate behaviour between two people in one situation is not necessarily appropriate in another situation.

VII. A COMPARATIVE STUDY ON FORMAL AND INFORMAL NETWORKS

The following table briefly describes about our review, observations and finding on contrast between formal and informal networks.

TABLE III
some contrast between Formal and Informal Organizations

Elements	Organization	
	Formal	Informal
Salient Goals	Organization's	Individual's
Structural Units	Offices/positions	Individual roles
Basis for communication	Offices formally related	Proximity: Physical, professional task, social, formal
Basis of power	Legitimate authority	Capacity to satisfy individuals (often through expert or reference power)
Control mechanisms	Rules	Norms
Type of Hierarchy	Vertical	Lateral
Belonging of Individuals	Specific	Ambiguous
Communications	Structured	Unstructured
Origin	Planned	Spontaneous
Changes over time	Shifts	Incremental
Group leadership	Explicit	Implicit

VIII. NEED FOR EXISTENCE OF INFORMAL NETWORKS

Affiliation needs: To satisfy the need for belonging to a group, individuals will tend to join networks of friendship and support. As a result, a part of one's individuality is sacrificed in order to conform to the group norms.

Identity and self-esteem: Conversely, belonging to a group or informal network can develop, enhance and confirm an individual's sense of identity as a result of the personal interaction.

Social reality: Since traditional formal organizations offer little room for emotions, feelings or sharing of personal thought, informal networks serve as an agent for structuring and supporting a shared social reality. By relying on this social reality, individuals can reduce uncertainty and stress.

Defence mechanism: In the face of a perceived threat or general uncertainty, group cohesion can act as a defence mechanism to reduce (perceived) uncertainty and strengthen each individual's ability to respond to that threat.

Risk reduction: Through diluting blame and aggregating praise, a group of individuals have a lesser perceived risk than they would have as individuals. Thus the unconscious efforts of individuals to control the conditions of their existence will lead to the creation of informal groups (Farris, 1979; Likert, 1961) [5][17]. Apart from those reasons, more practical and often conscious reasons for the creation and development of informal networks also exist:

Need to know: One of the primary characteristics of the informal structures in organizations is their communication network, often referred to as the grapevine. Studies have shown grapevine communication to be both fast and surprisingly accurate (Crampton & Hodge, 1998) [3], and in situations when information is critically needed by an individual to perform the task at hand, the grapevine can prove an efficient vehicle for news and information, thus bypassing the formal channels of communication (Mintzberg, 1975) [19].

Greasing the rusty wheels: Based on the same principles as the ‘need to know’-factor, individuals in organizations will tend to seek help from others and exchange favours to get things done, even when it entails stepping outside of the formal boundaries of the working units.

Political manoeuvring: One of the more conscious reasons for using informal networks, individuals might want to use the informal channels of communications to influence colleagues or superiors in order to gain an advantage in the organizational politics (Cobb, 1986; Pfeffer, 1981) [2] [24].

IX. A REAL LIFE EXAMPLE OF INFORMAL NETWORKS

We have already discussed about link analysis and literature related to it in earlier section. Informal networks is the situation, where there exist no defined or formal structure between nodes (people belonging to the network). It's more like creating an ad-hoc relationship.

Further considering the same scenario of institute environment and the example of director to student relationship as was discussed in previous section. There is no formal network exists between mentioned entities. As there is no existence of a formal network between the Director and the student and vice-versa the information can't be exchanged between them. The information needs to be delivered through the formal network which exists between the Director and HOD and then from HOD to the faculties to finally reach the students through the faculties. Similarly a student with any queries needs to go through the formal network in the reverse order as mentioned in order to reach the director. There is no formal network. But, still through link, the information flows, and it flows in both direction (i.e. from director to student and from student to director). Thus this is the case of Informal network and its application.

X. CONCLUSION

This paper starts with a formal introduction. We have divided this paper into two broad sections – namely, Link Analysis and Informal Network. Both are important part of social network analysis. The discussion on Link Analysis starts with an overview followed by its components. A real life example and major issues of Link Analysis have been discussed here. During our study of Informal Network we have studied its contents and the properties of the links in Informal Networks. A comparative analysis between Formal and Informal Network is mentioned here. Lastly, the literature of this paper concludes mentioning the need for existence of informal network and propositions concerning Informal Networks.

REFERENCES

- [1] Allen, T. H. “Communication Networks - The Hidden Organizational Chart”. The Personnel Administrator, 21(6), 31-35. Business School Press, 1976.
- [2] Cobb, A. T. “Informal Influence in the Formal Organization: Psychological and Situational Correlates”. Group and Organization Studies, 11(3), 229-253, 1986.
- [3] Crampton, S. M., & Hodge, J. W “The Informal Communication Network: Factors Influencing Grapevine Activity”. Public Personnel Management, 27(4), 569-584, 1998.
- [4] Etzioni, A. “Dual leadership in Complex Organizations”. American Sociological Review, 30(5), 688-698, 1965.
- [5] Farris, G. F. “The Informal Organization in Strategy Decision-Making”. International Studies of Management & Organization, 9(4), 37-62, 1979.
- [6] Goldberg, H. G. and Wong, R. W. H. “Restructuring Transactional Data for Link Analysis in the FinCEN AI System”, Proceedings of the 1998 AAAI Fall Symposium on Artificial Intelligence and Link Analysis, 1998.
- [7] Groat, M. “The Informal Organization: Ride the Headless Monster”. Management Accounting, 75(4), 40-42, 1997.
- [8] Han, P. E. “The Informal Organization You've Got to Live With”. Supervisory Management, 28(10), 25-28, 1983.
- [9] Hofstede, G “Cultural Constraints in Management Theories”. The Executive, 7(1), 81-40, 1993.
- [10] Hollingsworth, A. T. “Perceptual Accuracy of the Informal Organization as a Determinant of the Effectiveness of Formal Leaders”. Journal of Economics and Business, 27(1), 75-78, 1974.
- [11] Ibarra, H. “Structural Alignments, Individual Strategies, and Managerial Action: Elements Toward a Network Theory of Getting Things Done”, Networks and Organizations - Structure, Form, and Action (pp. 165-188). Boston, Massachusetts: Harvard, 1992.
- [12] Jensen, H. S. “Paradigms of the Theory-building in Business Studies”. In T. J. Elfring, Hans Siggaard; Money, Arthur (Ed.), European Research Paradigms in Business Studies (pp. 13-28). Copenhagen: Handelshøjskolens forlag, 1995.

- [13] Johanson, J.-E. "Formal Structure and Intraorganisational Networks". An Analysis in a Combined Social and Health Organisation in Finland. *Scandinavian Journal of Management*, 16, 249-267, 2000.
- [14] Krackhardt, D. "Graph Theoretical Dimensions of Informal Organizations". In K. M. P. Carley, Michael J (Ed.), *Computational Organization Theory* (pp. 89-111). Hillsdale, New Jersey - Hove, UK: Lawrence Erlbaum Associates, 1994.
- [15] Krackhardt, D., & Hanson, J. R. "Informal Networks: The Company Behind the Chart". *Harvard Business Review*, 71(4), 104-113, 1993.
- [16] Krackhardt, D., & Stern, R. N. "Informal Networks and Organizational Crisis". *Social Psychology Quarterly*, 51(2), 123-140, 1988.
- [17] Likert, R. *New Patterns of Management*.: McGraw-Hill, 1961.
- [18] "Link Analysis Workbench", Air Force Research Laboratory Information Directorate, Rome Research Site, Rome, New York, September 2004.
- [19] Mintzberg, H. "The Manager's Job: Folklore and Fact". *Harvard Business Review*, July-August, 49-61, 1975.
- [20] Mintzberg, H. "Structure in Fives". Englewood Cliffs, NJ: Prentice-Hall, Inc, 1983.
- [21] Mintzberg, H. H., Ludo Van der. "Organigraphs: Drawing How Companies Really Work". *Harvard Business Review*, 77(5), 87-94, 1999.
- [22] Monge, P. R., & Eisenberg, E. M. "Emergent Communication Networks". In F. M. P. Jablin, Linda L.; Roberts, Karlene H.; Porter, Lyman W. (Ed.), *Handbook of Organizational Communication* (pp. 304-342). Newbury Park, London, New Delhi: Sage Publications, 1987.
- [23] Newman, M. E. J. "The structure and function of complex networks", *SIAM Review*, Vol. 45, pp. 167-256, 2003.
- [24] Pfeffer, J. "Power in Organizations". Marchfield: Pitman Publishing, 1981.
- [25] Reif, W. E., & Monczka, R. M. e. a. "Perceptions of the Formal and the Informal Organizations: Objective Measurement Through the Semantic Differential Technique". *Academy of Management Journal*, 16(3), 389- 403, 1973.
- [26] Stephenson, K. *Social Networks*, 1998.
- [27] Tichy, N., & Tushman, M. "Social Network Analysis for Organizations". *The Academy of Management Review*, 4(4), 507-519, 1979.
- [28] Wasserman, S., Faust, K., and Iacobucci, D. "Social Network Analysis: Methods and Applications", Cambridge University Press, 1994.
- [29] Subrata Paul, Anirban Mitra & Prasanta Padhi, "On some aspect of Education Network", accepted for publication in *Oriental Journal of Computer Science and Technology*, June 2013.
- [30] Anirban Mitra, Sangita Rani Satpathy & Subrata Paul, " Clustering in Social Network using Covering Based Rough Set" *IEEE – Int. Conf. (IACC 2013)*(available in *IEEE Xplorer*) Ghaziabad, India, (10.1109/IadCC.2013.6514272), Pp. 476-481, Feb 22-23, 2013.