4th Construction Specialty Conference
4e Conférence spécialisée sur la construction

Montréal, Québec
May 29 to June 1, 2013 / *29 mai au 1 juin 2013*

Analyzing Infrastructure Discussion Networks: Order of "Influence" in Chaos of "Followers"

M. Nik Bakht, T. E. El-Diraby
Center for Civil Informatics, Department of Civil Engineering, University of Toronto

Abstract: This paper presents a summary of the tools that can be used to analyze public discussion over social media regarding infrastructure projects. It uses the cross-town transit project in Toronto as a case to illustrate how social network analysis can help decision makers to understand the composition and general trends of local community opinion. It focuses on analyzing the connectivity of the nodes and tries to benchmark best tools from network analysis to analyze the influence in infrastructure discussion networks. The tools are used to analyze micro-blogging website Twitter for cross-town to show how patterns of influence can be detected, and how they can be useful in decision making process. Analysis of the content (as the other component for understanding the social dialogue) is currently underway and would be discussed elsewhere.

1 Introduction

Modern urban infrastructure is a complex system composed of the physical network of the assets, social network of actors and users, and their interactions through the operational processes. This, as suggested by Lukszo & Bouwmas (2005) among others, redefines the civil infrastructure as a socio-technical system, rather than a mere technical/physical artifact. At the same time, social trends such as the proliferation of connectivity through Web 2.0 and the increasing awareness of sustainability are reshaping the public discourse on infrastructure. We must understand and accommodate the fact that the society cannot be considered any more as a background sound in the procedure of decision making (Parkin (1994)). A wealth of discussion is taking place online (instead of in community halls) and communities are becoming vocal in asking for a leading role in the decision making process (or at least in the idea generation). Day by day, decision making for development and operation of the modern civil infrastructure is evolving into a network-based system, where the public and their "public officials" along with professionals and media are all engaged in the debate. Web 2.0 provides the platform to support such a heterogeneous network which internally connects the actor network of decision makers into the community decision collaborators who – in form of end users – redesign the infrastructure continuously and dynamically from within the system (Kroes, et al. (2006)). We refer to such a large/heterogeneous network as '*Infrastructure Discussion Network*' (IDN). Bruijn & Heuvelhof (2000) confirm the inside-out characteristic of decisions made through the network, and define it as a 'process' rather than a 'project', which can be no longer initiated and controlled by a single (or even a single coalition of) actor(s). Decision making in this form will have a chaotic nature and role of official decision makers will change from "the initiator" into "a process architecture"/"process manager" to handle the chaos towards a final consensus.

Although reaching optimal consensus amongst technical and non-technical stakeholders in the above scenario seems nearly impossible, consultation and negotiation with the public decision contributors towards a solution with an acceptable level of satisfaction among the overall interests could be achieved by the official decision makers through the use of social connectivity. Social connectivity, specifically through the Web2.0 (social web) is an important enabler which makes such a goal accessible. The e-society not only has raised its voice through the transparent environment of the Web, but also can take the advantage of knowledge epidemiology to actively participate in the process of decision making for a technical issue – such as construction of civil infrastructure – in a self-organizing manner. As construction of the civil infrastructure becomes part of the bigger discussion about the sustainability and vitality of its community, the infrastructure discussion network allows community to actively participate in the debate as well as lead the decision making process. The socio-technical nature of infrastructure systems will require changes to organizational policies and rules. Decision process architectures must analyze and understand IDN and their contents to reshape project designs and plans.

Semantic analysis of the contents and evaluation of influence level of the speakers would be two central components for understanding and evaluating the IDN. This paper is focused on the second component. Social Network Analysis (SNA) tools can help the process manager in this regard by scratching the exterior layer of anarchy, detecting the order which exists beneath the chaos, and using it as a self-organizing evaluator for the nodes of the network. After a quick review on the previous works on SNA in construction, we try to makes the case for existence of infrastructure discussion networks over the online social media. Then we benchmark some tools from social network analysis and apply them to the IDN of an actual project to crystalize different patterns of influence which exist underneath the randomness of the complex network.

2 Related Works

A ‘network’ at the highest level can be defined as a set of interdependent individuals (nodes or vertices) with all of the interactions among them (edges or links). ‘Social Network Analysis’ is concerned with understanding the social linkages (relational ties) among the social entities (actors) as well as implication of these linkages. This involves measuring and studying structural, composition, or affiliation attributes/variables of a set of actors. While ‘Structural variables’ measure ties of a specific kind between pairs of actors, ‘Composition variables’ are measurement of actors’ attributes. ‘Affiliation variables’ are related to the relationships among the actors and ‘events’ which the actors belong to. Among all different structural properties of networks, research in SNA is primarily focused on diameter (the maximum length of shortest paths, or the longest social distance between a pair of nodes in the network), degree distribution (probability distribution for degree of the randomly selected node), and clustering coefficient (portion of a node’s neighbors who are connected to each other). These parameters reflect some important aspects related to formation and evolution of the social networks. For example, the effect of ‘popularity’ in social networks causes the degree distribution of such networks to follow the power-law[1]. This is a well-known behavior in various types of networks wherever there exist many of nodes with low degrees and just a few nodes with high degrees. Networks following power law are called “scale-free” networks. In social networks, this behavior is a direct result of existence of the feedback loops.

Different indicators are used in network science to examine characteristics of nodes and links. ‘*Centrality*’ is an indicator which gives a measure of importance for a node in a network. Different forms of centrality are defined in graph theory and each looks at a different aspect of importance of a node. The simplest form is ‘*degree centrality*’, which is equal to the degree (total number of neighbors) of the node. ‘*Closeness centrality*’ of a node is defined as number of steps required to access all other vertices from the given node and refers to accessibility of nodes in a graph. ‘*Betweenness centrality*’ of a node is the total number of shortest paths in a graph which pass through that node and therefore, a node of high betweenness centrality will be in a strategic position to control the connections among other individuals. ‘e*igenvector centrality*’ relates to being connected to high scoring nodes of the graph (such as nodes with high degree), and is calculated from the eigenvector of the adjacency matrix of the graph.

Social networks are mostly composed of groups of nodes which are tightly connected among themselves and are sparsely connected to nodes from other groups. These densely connected cores are called

[1] Sometimes referred to as: long tail, heavy tail, or Zipf’s law

‘*clusters*’, ‘*modules*’, or ‘*communities*’. This behavior has roots in the formation process of the network and is due to the fact that people are more likely to join the communities in which not only they have more friends, but also the friends are more densely connected to each other (Social capital argument by Coleman (1988)). ‘Modularity maximization’ is one of the most popular methods for partitioning graphs of social networks into their communities. Modularity is defined as the difference between number of existing edges in a partition and expected number of edges which can exist among the nodes of that partition. The higher modularity of a partition is, the denser that community will be. Therefore, problem of community detection can be formulated as partitioning a graph into groups such that the summation of modularity over all partitions is maximum. Various methods are suggested in literature for solving such a maximization problem and for a comprehensive review on some of them, one can look at Fortunato (2010). *Multi-level aggregation* algorithm by Blonder, et al. (2008) is the one used in this paper. This is a simple heuristic optimization technique working based on forming small communities by local optimization of modularity first, and then aggregating nodes of the same community and forming bigger communities.

Review on the literature of social network analysis in construction shows that studies are mostly focused on the ‘*project network*’, which is composed of technical actors, together with the interactions they have in order to deliver a project. Such networks are small sized in scale of the networks commonly being studied in analysis of online social networks. In fact, construction projects have been modeled as network of stakeholders exchanging information, as early as 1950’s (Stinchcombe (1959)). Taylor and Levitt (2007) define the project network as “a group of specialist firms contracted to work together on specific construction projects”. Pryke (2004) also looks at the construction project as a network of interactions, interdependencies, and information exchange, and argues how a network perspective of the construction project can be more mature than the traditional hierarchical management structure. Over the past decades, globalization has fostered the phenomenon of networked organizations which are sometimes spread all around the globe (Wong, et al. (2010)). Enablers such as the Web and cloud computing have also been supporting this trend. More recently, researchers proposed that the interactions among project staff is a form of social network. Chinowsky, et al. (2008) propsoed a model for construction projects as a social network having two main components, ‘*dynamics’*: to address the dynamics of interactions among the actors, and *‘mechanics’:* to model the free flow of knowledge among project participants. This model is validated in some empirical case studies by Chinowsky, et al. (2010). Identification of relationships among the actors, tracking flow of information, and transparency of feedback loops are mentioned among other necessary criteria for success of the construction project network by Boddy, et al. (2010). Published research in construction is also partly focused on applications of structural SNA tools for analysis of networked organizations in construction industry. ‘Centrality’ has been the most popular indicator used in the domain with the intention of measuring degree of importance for nodes and edges of networks. Table **1** summarizes some of the applications of centrality in analysis of importance in project networks.

Table 1: Centrality as used in construction for analysis of importance in the project network

Reference	**Effect Studied**	**Degree Centrality**	**Betweenness Centrality**	**Closeness Centrality**	**Eigenvector Centrality**
Loosemore (1994)	Crisis control in construction project	✓	✓	✓	
Thrope & Mead (2001)	Efficiency of the project-specific website in construction projects	✓			
Pryke (2004)	Project governance	✓			
Di Marco, et al. (2010)	Cross cultural barriers and collaboration effectiveness	✓	✓	✓	
Park, et al. (2011)	Collaborative venture for international projects	✓	✓		
Wambeke, et al. (2012)	Key trades in construction project with multiple trades	✓			✓

3 Infrastructure Decision Network (IDN)

As the social dimension of infrastructure projects becomes more and more central, social media tools (such as blogs, Facebook and twitter) are being used more progressive to enable better involvement of the community in projects. General public is able to express opinions and share ideas amongst themselves and with professional staff or city officials. As these non-technical users of the project are added to the project network, not only the contents of the discussion changes, but also the complexity of the network increases. On one hand, managing such a complex/non-technical interaction burdens decision makers with a challenge; on the other hand, such networks present a rich source for understanding community needs and interacting with the community members to explain/discuss project options and features. As diffusion of information is one of the core functionalities of online social media, it can add a remarkable value by completing a transparent path for bidirectional, real-time flow of information from official decision makers to unofficial contributors and vice versa.

Social network analysis has been used by researchers and industry alike to manage a variety of business needs-such as user profiling, customer relationship management, and reverse marketing. Two main approaches have been used in this regard: text mining/analysis and structural network analysis. By analyzing the content of stakeholders' inputs we can detect the main concerns and needs of the target community along with their trends over the time. Further, by analyzing the social interrelationships between the individuals we can detect sub-communities and also study the progression of ideas. One great example of an online social medium which provides us with such an opportunity is the micro-blogging website Twitter. It not only archives the social opinions, but also keeps the record of connectivity for individuals. Therefore, 'networkedness' of the ideas can be distilled by studying Twitter. Although Twitter network might not be an exact representative of the real – offline or even online – social life; as suggested by Huberman, et al. (2009), the one-sided relationships on Twitter defined at different levels[1] can be adequately taken as indicators of the social influence. In the next part of this paper, network of followers of the twitter account for an LRT project will be studied as the example of an IDN.

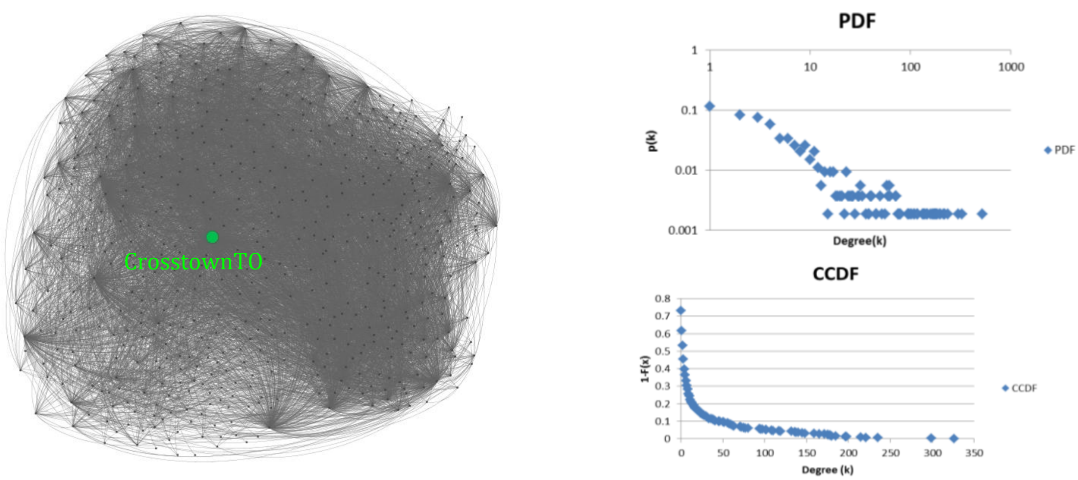
Various semantic methods for text analysis, opinion mining, and sentiment analysis exist to help understand and classify the content of discussions over the IDN. However, as the topics are being discussed over a network, the ideas must be evaluated with respect to their proponents, and based on the level of influence the person who is discussing them. But evaluation of influence degree for a node in the IDN would require more than just reading attributes such as followers count and number of tweets from profile of that individual. As IDN is an issue-centered sub-network of the general social network, it must be cut out first, and then the internal patterns of influence in it should be detected and quantified. Centrality – as discussed earlier – is the indicator which traditionally has been used as a measure of importance in the project network (see Table 1). However, different nature of IDN – in terms of size, heterogeneity, and complexity – requires the applicability and efficiency of this tool to be re-examined. Degree centrality only counts the number of neighbors of a node, and misses their connectivity and level of importance. Closeness centrality and betweenness centrality on the other hand, are respectively measures for accessibility and strategic behavior of nodes and are less relevant here. In order to examine level of influence of nodes in IDN, not only the *quantity* (number of people who are under influence of an individual – called followers count) matters, but also the *quality* (level of importance) of the followers must be considered. Although eigenvector centrality is meant to give such a measure, it still takes the number of neighbors' followers (quantity) as their importance factor, and ignores the degree of importance (quality) of the followers. There might always exist some nodes with few-and yet high-influential followers. This will give leverage to such nodes, and makes them top influential individuals of their network. More advanced tools are developed in the literature to analyze the influence from this perspective. In this paper, patterns of influence in the IDN will be detected by applying methods benchmarked form webpage ranking and SNA. This is a basic example to address how the influence can be evaluated among nodes of the IDN, and how this can be helpful in analysis of structure of social composition, formed in online media around development of urban infrastructure.

---

[1] Ranging from weak ties of followership to more close connections such as re-tweeting, mentioning, or even direct messaging

4 Case Study

The Eglinton-Scarborough "Crosstown" (a light-rail system) is one of the largest transit projects currently underway in North America. It is part of a bigger city-wide transit plan called "Transit City", which was announced in 2007 and has been under long debates since then. Cancellation ofTransit City by the mayor in the late 2010 and resuming it again in early 2012 by Toronto City Council under several causes such as pressure of the community (mainly in form of a social movement called: 'Save Transit City'), are among other reasons which put this project under spotlight of social attention. Crosstown is an $8.2 Billion, 25.2 kilometer east-west LRT line passing through a congested corridor of the midtown of Toronto and is running underground in major parts (19.5Km). The street-level segment is planned to be separate from the street traffic with raised medians. TTC (Toronto Transit Commission) is the eventual owner and operator of the project. Metrolinx (a Provincial planning and finance agency) is the owner's agent in construction. Several Canadian contractors and consultants are the other "technical" partners of the project. Procurement began on March 2011with manufacturing of the pre-cast tunnel linings and the opening is planned for 2020. The construction officially launched on November 2011, and the tunneling is expected to begin by summer 2013. Considering the history of urban transit plans in the city[1] which proves the high cost of social opposition, as well as the high sensitivity of the community in the city of Toronto regarding transit projects, several community meetings for public consultations have been held and more are planned on different aspects of the project (such as general specifications, station designs, construction schedule, and operation plan). In the era of online social media, official decision makers also launched a Twitter profile for the project on December 2011. A year later, it has more than 700 followers.



(a) IDN for the Crosstown project on Twitter (b) Degree distribution of the IDN follows power law

Figure 1: Infrastructure Discussion Network as an ego–centered graph, and its degree distribution

Connectivity among followers of the Crosstown project has been detected by collecting data from Twitter API (Application Programming Interface). The result, as of August 2012, is the graph shown in Figure **1**(a), where each node represents a Twitter profile and a directed edge from node A to node B indicates that A is following B on Twitter. Version 0.8.1 of Software *Gephi (an open source software for network visualization: https://gephi.org/)* is used to visualize the network. This is an 'ego-centered' network with 'CrosstownTo' as its focal actor (termed ego). Ego-centered networks consist of a focal actor and a set of alters who have ties to it. Such networks are widely used by anthropologists to study the social environment surrounding individuals or families, as well as by sociologists to study the social support. Collecting data for users with protected profiles is not possible. However, as the main focus of this study is 'influence', it can be simply assumed that while tweets by such individuals are not publicly available; these actors cannot have a high degree of impact on the network. Therefore, such profiles are nodes of

[1] Case of Spadina expressway as discussed by Frisken (1988) among others, and case of St.Clair West transit improvement as discussed in TTC (2010) can be mentioned as examples.

the graph with only one outgoing link (termed 'orphans') following the project profile only. Some filtering is now needed to crystalize the islands of order existing in the sea of chaos[1] of the amorphous bulk of nodes and links seen in Figure 1(a). We use two types of filtering here to identify the top influential nodes (which sometimes are hidden in the chaos of community participation), and to detect the clusters of users following the project.

4.1 Detection of Top Influential Actors

Ego-centered network of CrosstownTo followers totally had 522 nodes and 9535 links (i.e. the average degree of 18.2) when we detected it on August 2012. Average path length of this network (as a directed graph) is 2.45, and diameter of the directed graph is 6. Nodes' degrees in the graph (in-degree, out-degree, and total degree) follow a power law distribution as shown in Figure 1(b). Average clustering coefficient of the network is 0.68. This means that on average, almost 70% of neighbors of a node are connected to each other. These topological specifications more or less verify that the network of crosstown project followers has the general behavior of a social network[2]. In order to investigate the influence, we start with analysis of centrality (as the most frequently used indicator in the literature of construction). The most naïve method to find nodes with high level of influence is looking at the degree centrality of nodes. Having a directed network of followership, in-degree (number of the node's followers over IDN) will be a measure of popularity and can somehow imply the influence on others. Looking at the profile of the nodes with high degree centrality shows a mixture of actors with different roles regarding the project: Official decision makers (including city councilors, mayor of Toronto, and the project owner-TTC chair), technical decision makers[3] (including some management level of the TTC, as the project owner, as well as the general contractor), and finally, public decision contributors (mostly composed of journalists, reporters, social activists, urban planners, etc.). This mixture of roles verifies the heterogeneity of the nodes contributing to the infrastructure discussion. However, ranking nodes in a self-organizing/networked manner must be emanated from their degree of impact they have on others, regardless of their role. Other types of centrality are also tested here. As scope of this study is social influence, closeness centrality would not be relevant. Analysis of betweenness centrality shows that the same nodes with high degree centrality more or less have high betweenness centrality (slightly different in the order).

In order to measure being connected to high scored nodes, one may look at the ranks based on eigenvector centrality. However, In this paper, we use HITS[4] analysis and page ranking methods for this purpose to take both quality and quantity of the followers into account. These methods can also distinguish between the dual role of nodes in an IDN, which will be discussed later. In directed graphs, importance of a node can be generally measured from two different aspects. The concept originally comes from Webpage ranking, where '*authorities*' are pages containing quality information (and therefore receiving lots of links from several '*hubs*'), and hubs are pages which link to the top authorities. As it is seen, the two concepts are interdependent and therefore, calculating these measures will be a recursive procedure. Kleinberg (1999) algorithm is used here for this purpose. This algorithm assigns a non-negative *authority weight* and a non-negative *hub weight* to each node. These factors are normalized such that summation over all nodes, for each factor adds up to one. The algorithm calculates hub weight for each node as the summation of the authority weights of all the nodes it directs to, and the authority weight of each node as the summation of the hub weights for all the nodes which point into it. The weights are assigned this way and get updated recursively till they converge. Nodes of IDN with high authority will be individuals of the network who are followed by a large number of 'active' followers. As the network is a directed-connected graph, authority will exactly lead to the same ranking as the in-degree (see Table 2). Looking at the hub weights however, shows that except from the project itself (which is not

[1] Term used by well-known chaos theorist mathematician Ian Stewart

[2] As shown by (Kwak, et al., 2010), the global graph of Twitter does not follow some general behaviors of social networks (such as power law degree distribution and high clustering).

[3] Note that by technical decision maker, we are referring to the 'role' of an individual with respect to this specific project, not only an actor with a technical background

[4] Hyperlink-Induced Topic Search

a huge hub), the other authorities are [with a slight difference in order] at the same time main hubs of the network, (for the first top five, even the order is not different). This may suggest that in this project, there is a relatively close mutual connectivity among top nodes of the network. However, this is not the case when we look at the nodes at the lower levels of influence ranking.

The HITS analysis is based on two main assumptions which transfer from Webpage ranking: first, nodes play two different roles in the network; and second, good authorities receive links from good hubs, and top hubs link to good authorities. Although the first assumption applies to the IDN[1], the second one has to be revised. An influential node is not necessarily the one being followed by those who are following lots of other people; rather, we can think of a node which is followed by most popular nodes of IDN (the ones with lots of followers) as an influential actor. In simple words we are looking for authorities followed by other good authorities. In order to detect such nodes in our network, we use one of the most popular page ranking algorithms, called ‘PageRank’. This algorithm, as introduced by Brin & Page (1998) starts by allocating equal weights (called PageRank) to all nodes and works based on flow of weight among nodes of the graph. It is another recursive algorithm in which nodes transfer their weights to their neighbors whom they are following, and weights of nodes get updated over a preset number of iterations. As the result, weight of a node is defined with respect to its neighbors and based on number of links it is receiving from other individuals with high degree of influence. Figure 2 shows a scaled view of the network based on this ranking method. Nodes of higher PageRank have bigger size in this figure.

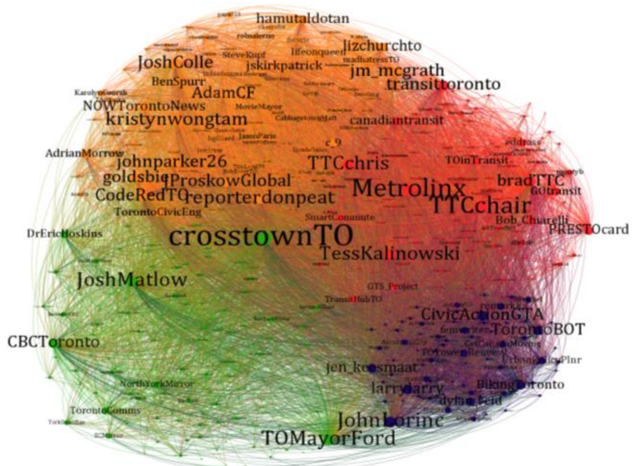


Figure 2: Communities of ‘CrosstownTo’ followers (size of a node reflects its level of influence based on PageRank and color of a node shows the community it belongs to)

Table 2 gives the top influential actors based on the PageRank, and compares it versus other methods. The owner/operator (TTC), as the top official decision maker of the project ranks first in four out of seven methods. Metrolinx (Financier and Owner’s Agent) is the second top node. Position of mayor of Toronto in this table is noteworthy. Opposing Transit city as one of the central themes of his election campaign ties his name closely to the name of this project, and that is why in almost all different types of influence ranking, his name is among the top few nodes. In terms of comparing centrality versus page ranking methods, as it is seen, more or less the same nodes are detected as the top actors by all techniques (except closeness centrality), though the order is different for different methods. However, when it comes to the lower levels of influence ranks, there are differences between evaluations given by centrality and PageRank. This is due to the fact that the former somehow ignores the degree of importance of followers. Another observation is that followers count (degree of a node on the general social network: Twitter) does not necessarily imply the level of influence of that node within the IDN. For example, the current network –being ego-centered– has CrosstownTo –the project profile/focal point– as the node with the highest number of followers (in-degree centrality) in the IDN; however, in terms of followers count on the Twitter,

[1] in process of decision making, an individual in this network can add value in two different ways: providing information (hubs), or influencing others (authorities)

CrosstownTo is ranked 112 among its own followers! On the other hand, while the highest number of Twitter followers count belongs to 'CBCToronto' (an important news media with 38,563 followers over Twitter), it is ranked 15th based on the centrality within the IDN. This is due to the fact that IDN is an issue centered social network where people with a particular common interest (here: a transit project) get together.

Table 2: Ranking top influential nodes based on page ranking and centrality

Node ID	Description	Followers Count	Rank						
			Page Ranking			Centrality			
			PR	A	H	C_D	C_B	C_C†	C_E
CrosstownTo	*Profile of the project*	521	0	0	521	0	360	521	0
Metrolinx	*General Contractor*	3778	1	2	2	2	5	433	2
TTCchair	*Owner*	4850	2	1	1	1	6	410	1
JohnLorinc	*Journalist, reporter*	3557	3	5	5	5	25	327	3
TOMayorFord	*Mayor of Toronto*	23414	4	4	4	4	9	311	9
JoshMatlow	*City councilor*	5199	5	3	3	3	16	397	6
TessKalinowski	*Transit Reporter*	1568	6	11	9	11	8	431	5
TTCchris	*Owner's technical*	1863	7	6	7	6	3	485	10
reporterdonpeat	*Journalist, reporter*	4089	8	8	8	8	10	448	4
kristynwongtam	*City councillor*	6405	9	7	6	7	12	469	8
JoshColle	*City councillor*	2291	10	14	14	14	11	421	11
jm_mcgrath	*City activist*	3397	11	16	16	16	15	470	7

PR: PageRank **A**: Authority **H**:Hub **C_D**: In-Degree Centrality

C_B: Betweenness Centrality **C_C**: Closeness Centrality **C_E**: Eigenvector Centrality

† Closeness centrality is irrelevant in our study but is listed here to complete the comparison of page ranking versus all popular forms of centrality used in previous research in construction social network analysis.

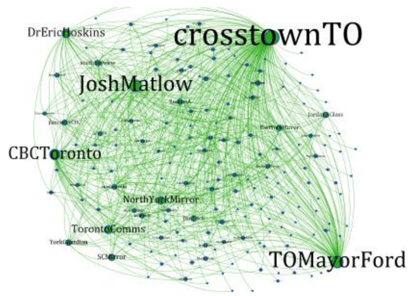
4.2 Patterns of Influence within the Communities

Applying Multi-level aggregation algorithm to the network of CrosstownTo detects four communities which are highlighted with different colors in Figure 2[1]. The communities are numbered here based on their size. This numbering strategy also complies with the order of degree centrality of the groups[2]. Gephi Software is used again for coloring and visualizing communities. When we apply the influence measuring filter to the clustered graph, relations between top nodes of a community with each other, and with important nodes in other communities are revealed. Focusing on each community in isolation, ranking its nodes based on their degree of influence, and detecting their identities from their profiles on the social media result in a more profound understanding of the social composition and the way society has structured itself around the project. Figure 3 illustrates the four communities detected in the 'CrosstownTo' IDN. The first community (Figure 3(a)) is under influence of Toronto mayor, news media, and some other politicians such as some provincial ministers and members of provincial parliament (MPPs). Top influential nodes of the second community are mostly technical actors of the project (Figure 3(b)). Owner and their technical consultants (head of TTC and TTC technical managers), general contractor, and some of the organizations which are created or being run by them have the highest weights in this cluster. Being a governmental project, apparently some of the official decision makers (such as the provincial minister of infrastructure and transportation) can also be found in this cluster. Community3 is mostly dominated by city councillors, as well as City Hall reporters and columnists who cover the news related to them. Organizations working under the city government (such as public consultations and civic engagement) also belong to this cluster. Finally, the last community is mostly under influence of a combination of public followers of the project: urban planners, some urban strategists, transit designers and transportation engineers (without any official affiliation to this project). Some social activists who follow the

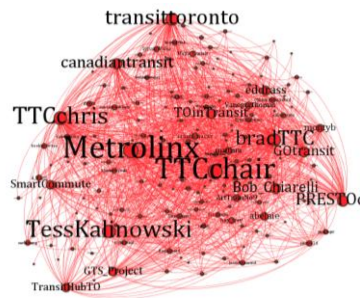
[1] It must be noted that this algorithm works based on a random process and each time it is applied, results might be slightly different. However, what is presented here is one of decompositions with the highest total modularity.

[2] Degree centrality of a subgraph is defined as the summation of differences between the highest degree in that subgraph and node degrees over all nodes of the subgraph.

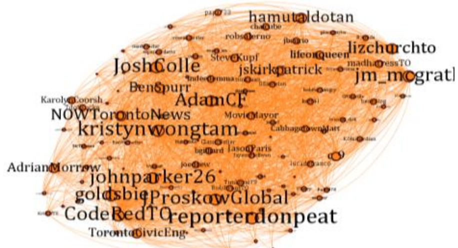
transportation plans in Toronto (such as organizers and coordinators of the social movement 'Save Transit City') are also among nodes of this community.



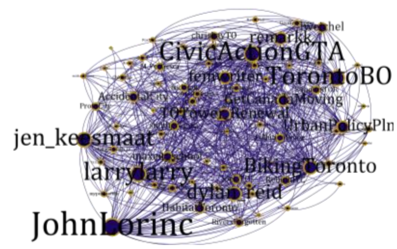
(a) Community1, 32.95% of total nodes – Top influential nodes: Political



(b) Community2, 30.27% of total nodes – Top influential nodes: Technical/Official Decision makers



(c) Community3, 22.03% of total nodes – Top influential nodes: Policy Makers



(d) Community4, 14.75% of total nodes – Top influential nodes: Public

Figure 3: Communities under microscope; ranking community nodes by PageRank

Results of the detailed analysis of communities can be summarized as follows:

- Using the two filters (community detection at the same time with the PageRank) is not only helpful, but somehow necessary to make the detected patterns meaningful.
- Although each group is mixed and composed of actors with different types of interests, and a clear/formal cut might not exist between different communities, by looking at the profile of individuals, one can clearly detect similarities in the role of top influential nodes of each community, with respect to the project.
- In communities 3 and 4 (unlike communities 1 and 2), differences between the nodes' degree of importance are lower. This can be due to the fact that the hierarchical structure of project's technical and official stakeholders does not exist among the public, and the city councillors.
- In this pubic community, community of the public has the lowest size and degree centrality. This can be justified based on the fact that the project is not at the construction phase yet. Public attention is usually attracted at the more visible phases of a project (i.e. construction).
- The term 'Public', here, refers to a collection of nodes with no official role. As we see, some of these nodes are experts who can have intuitive inputs for the official decision makers.
- Communities of the IDN are in fact the cores of interest among the participants and followers of the project. These are the main parties whose interests must be addressed by the package deal which will be suggested by the network of decision makers.

5 Discussion and Conclusion

Development of infrastructure as a sociotechnical system requires a wider range of non-official/non-technical decision contributors to be involved in the process of decision making in form of an Infrastructure Discussion Network. Such a network will have a different nature from the project social network in terms of size, heterogeneity, and complexity. In order to detect degree of influence for actors in such a network, popularly used measures such as centrality can be replaced by more sophisticated indicators. Algorithms used in the literature for webpage ranking (such as HITS analysis and PageRank) were benchmarked in this paper and it was shown how they not only suggest a more solid computational approach (which measures the degree of influence based on degree of influence of followers in an iterative manner), but also distinguish between dual roles of nodes in the IDN (information propagation–Hubs; versus influencing others' opinion–nodes with high PageRank value). It was also shown how using more than one filter might be necessary to detect the patterns of order in the chaotic environment of IDN. Many nodes from the public emerge after detecting community and searching within each community for the influence level.

Analyses similar to what was presented here can generate useful guidelines showing the structure of parties who must be involved in the negotiation process. Hubs of the network can be helpful if being involved in the initial phases such as requirements identification and preplanning. On the other hand, nodes with high PageRank value are the top influential actors and most probably are useful to be involved in the later phases when a suggestion is prepared and feedback is required, or in the process of negotiations for implementing the plan. In combination with a proper content analysis of the online discussions, influence degree of a follower can be used as a measure to evaluate weight for the topics discussed by that individual. Such an evaluation method has the eminence of being self-organizing and emerging out of interactions of the actors from within the system. Real interactions in the offline world are more or less reflected through the online behavior of the actors. This can provide a powerful tool for analyzing and guiding the communication process with the public. Since offline social opposition most of the time lags the online declaration of dissatisfaction, detecting online alarms will give the official decision makers enough time to change the decisions appropriately, or to apply timely policies to prevent formation of snowballing social opposition and to reduce the risk of failure in such projects before it is too late.

Finally, it must be noticed that Twitter (or any other social media) does not necessarily represent an exact picture of society. There will always exist people who have a high impact on their communities but are not active in online environment or do not have a Twitter account. Furthermore, the online, and offline social attitudes do not necessarily match perfectly all the time; for instance, people tend to vent extreme emotions online while they are generally more conserved in real life. Hence, existence of a followership relation between two nodes on the Twitter might not essentially prove that the followee has a firm influence on the follower. The ongoing research is focused on closer types of connection and stronger ties among the followers of projects (such as mentioning and re-tweeting). Study is also underway not only to examine the structure of the networks, but also to understand/cluster contents discussed over the IDN, and to analyze the interrelationship between the contents and connections. While the idea of combined network of official and non-official decision makers is a futuristic view which needs more experiments over the time to settle and stabilize, what was presented in this paper is applicable into any other platform in which connectivity of people discussing construction of infrastructure is recorded, and it has the notable advantage of highlighting hidden nodes and amplifying sounds which are regularly difficult to hear.

References

Blonder, V. D., Guillaume, J.-L., Lambiotte, R., & Lefebvre, E. (2008). Fast unfolding of communities in large networks. *Kournal of statistical mechanics: Theory and experiment, 10*, 2000-2012.

Boddy, S., Rezgui, Y., Cooper, G., & Wetherill, M. (2010). Activity Awareness as an Enabler for Communication and Network Building in Construction Design Teams. *Journal of Computing in Civil Engineering, 24*(5), 430-440.

Brin, S., & Page, L. (1998). The anatomy of a large-scale hypertextual Web search engine. *WWW7 proceedings of the seventh international conference on World Wide Web* (pp. 107-117). Elsevier.

Bruijn, H. d., & Heuvelhof , E. t. (2000). *Networks and decision making.* Utrecht: LEMMA Publishers.

Chinowsky, P., Diekmann, J., & Galotti, V. (2008). Social network model of construction. *ASCE Journal of construction engineering & management*.

Chinowsky, P., Diekmann, J., & O'Brien, J. (2010). Project organizations as social networks. *Journal of construction engineering and management, 136*(SPECIAL ISSUE: Governance and leadership challenges of global construction), 452-458.

Di Marco, M. K., Taylor, J. E., & Alin, P. (2010). Emergence and role of cultural boundary spanners in global engineering project networks. *Journal of Management in engineering, 26*(3), 123-132.

Frisken, F. (1988). *City policy-making in theory and practice: The case of Toronto's downtown plan.* London, Ontario, Canada: University of Western Ontario.

Huberman, B. A., Romero, D. M., & Wu, F. (2009). Social networks that matter: Twitter under the microscope. *First Monday*.

Kleinberg, J. M. (1999). Authoritative sources in a hyperlinked environment. *Journal of the ACDM, 46*(5), 604-632.

Kroes, P., Franseen, M., Van de Poel, I., & Ottens, M. (2006). treating socio-technical systems as engineering systems: sme conceptual problems. *System research adn behavioral science, 23*, 803-514.

Kwak, H., Lee, C., Park, H., & Moon, S. (2010). What is Twitter, a Social Network or a News Media? *The international World Wide Web conference .* Raleig, North Carolina: IW3C2.

Loosemore, M. (1994). Social network analysis: using a quantitative tool within an interpretative context to explire the management of construction crises. *Wngineering, construction and architectural management, 5*(4), 315-326.

Lukszo, Z. V., & Bouwmas, I. (2005). Inteligent complexity in networked infrastructure. *IEEE international conference on Systems, Man and Cybernetics,* (pp. 2378-2383). IEEE.

Park, H., Han, S. H., Rojaz, E., Son, J., & Jung, W. (2011). Social network analysis of collaborative ventures for overseas construction projects. *Journal of construction engineering and management, 137*(5), 344-355.

Parkin, J. (1994). A power model of urban infrastructure decision making. *Ceoforum*, 203-211.

Pryke, S. D. (2004). Analysing construction project coalitions: exploring the application of social network analysis. *Construction management and economics, 22*(8), 787-797.

Stinchcombe, A. (1959). Bureaucratic and craft administration of production: A comparative study. *dministrative Science Quarterly, 4*(2), 168-187.

Taylor , J., & Levitt , R. (2007). Innovation alignment and project network dynamics: an integrative model for change. *Project management journal*.

Thrope, T., & Mead, S. (2001). Project-specific web sites: friends or foe? *Journal of construction engineering and management, 127*(5), 406-413.

TTC. (2010). *Transit city implementation- the St.Clair project experience.* Toronto: Toronto Transit Commission.

Wambeke, B. W., Liu, M., & Hsiang, S. M. (2012). Using Pajek and centrality analysis to identify a social network of construction trades. *ASCE Jounral of construction engineering and management, 138*(10), 1192-1201.

Wong, K., Unsal, H., Taylor, J. E., & Levitt, R. E. (2010). Global dimension of robust project network design. *Journal of construction engineering and management, 136*(4), 442-451.