A Social Network Matrix for Implicit and Explicit Social Network Plates

Wei Zhou[1,4], Wenjing Duan[2], Selwyn Piramuthu[3,4]
[1]Information & Operations Management, ESCP Europe, Paris, France
[2]Information Systems and Technology Management, George Washington University, U.S.A.
[3]Information Systems and Operations Management, University of Florida, U.S.A.
Gainesville, Florida 32611-7169, USA
[4]RFID European Lab, Paris, France.
wzhou@escpeurope.eu, wduan@gwu.edu, selwyn@ufl.edu

Abstract

A majority of social network research deals with explicitly formed social networks. Although only rarely acknowledged for its existence, we believe that implicit social networks play a significant role in the overall dynamics of social networks. We propose a framework to evaluate the dynamics and characteristics of a set of explicit and associated implicit social networks. Specifically, we propose a social network matrix to measure the implicit relationships among the entities in various social networks. We also derive several indicators to characterize the dynamics in online social networks. We proceed by incorporating implicit social networks in a traditional network flow context to evaluate key network performance indicators such as the lowest communication cost, maximum information flow, and the budgetary constraints.

Keywords: Implicit Social Network, Online Social Network

1 Introduction

In recent years, several online social network platforms have witnessed huge public attention from social and financial perspectives. However, there are different facets to this interest. Facebook, for example, has gained a large set of users but failed to excel on profitability.

Online social networks can be broadly classified as *explicit* and *implicit* social networks. Explicit social networks (e.g., Facebook, LinkedIn, Twitter, and MySpace) are where the users define the network by explicitly connecting with other users, possibly, but not necessarily, based on shared interests. Implicit social networks (e.g., Last.FM, Outbrain, and Color) are networks where a user is defined by his or her interests and the (implicit) connections between users are not explicitly created by the users themselves but evolve purely based on their interests as exemplified by their online behavior. An *implicit* social network could be ephemeral and last only as long as is necessary, unlike a majority of explicitly created networks. For example, *Color* has the ability to co-locate users and determine their implicit social graph, that can then be used to introduce items from users who do not necessarily know one another. *Color* lets users who took and posted photographs from an event (e.g., wedding, game, music) to view photographs taken by other users from the same event using location-based metrics. Unlike explicit social networks, their implicit counterpart is not limited by users who are friends or acquaintances.

We investigate online implicit social networks and their unique characteristics when considered along with their explicit social network counterparts. We propose a measurement matrix that can be used to evaluate social networks not only by their topology, but also from a network flow perspective. With the social network matrix, practitioners can readily determine the key performance indicators of a single or a set of social networks.

The remainder of this paper is structured as follows. We first review existing related literature on social network measurement and online social network applications. We then present our social network matrix methodology in Section 3. In Section 4, we model the implicit social network matrix within the framework of traditional network flow context and evaluate three practical social network management problems, including the lowest cost problem, the maximum information flow problem and the budgeting problem. In conclusion (Section 5), we present our main findings and propose future avenues for research.

2 Related Literature

Explicit online social networks have been extensively studied by researchers during the last decade. However, implicit online social networks have not received their fair share of attention, possibly due to the difficulty in extracting them from readily available data. Nevertheless, the last few years have witnessed increasing interest among both researchers and practitioners to seriously consider implicit social networks/graphs. We now briefly discuss existing literature on implicit social networks.

As discussed in Wasserman and Faust (1994) and Lattanzi and Sivakumar (2009), social networks can be represented as an "affiliation network" in an organization, such that a publisher can reach its audience by promoting programs to its affiliation networks. The modeling approach in our paper shares some common traits with the traditional affiliate network perspective in that both models utilize the similarity between a business entity and the social network. In addition to a pure value-based analysis perspective, our model considers additional dimensions through incorporation of non-value (implicit) traits of a social network that is more fundamental to network analysis. Our model also differs from affiliate network models with the formation of relationships between any two affiliated social networks.

Nauerz and Groh (2008) consider the determination of expert users in a Web portal. To accomplish this, they stress the importance of understanding the users' behavior, their interests, preferences and knowledge. They use both static information from users' profiles such as their age and native language as well as dynamic information such as those that are retrieved from Web usage mining, user tag behavior, among others. Such implicit online social network information is then complemented by explicit online social network information to help determine expert users.

Smith et al. (2009) show how to generate individual-centered social networks which are not built around explicitly announced relationships, that they call 'implicit affinity networks (IAN),' which is another name for implicit social networks. These IANs capture dynamic, multi-faceted relationships that are implicit in the shared characteristics or attributes of individuals. They determine an individual user's social capital based on a hybrid network that comprises both the implicit and explicit form of online social networks using a mathematical formulation. In doing this, they decouple bonding and bridging social capital so that they are allowed to vary independently of each other.

Using online implicit social networks that is formed by a weighted graph with edge weights determined by the frequency, recency, and direction of interactions between users and their contacts and groups, Roth et al. (2011) present a friend-suggestion algorithm. They follow related literature in distinguishing implicit online social networks from online explicit social networks that are explicitly generated by the users themselves. This algorithm assists users to implicitly or explicitly create customized contact groups. They use interactions between users and their contacts as well as groups of contacts to generate the implicit online social network graph, which is then analyzed to operationalize the proposed algorithm. Their algorithm incorporates both group interactions and peer-to-peer interactions to determine tie strengths in the developed networks. As initial seed input to their algorithm, they use the user's social network with weighted edges and a sample from the user's contacts to generate a customized contact group that expands the initial seed of a few contacts. They illustrate their algorithm using implemented Gmail features "Don't forget Bob!" and "Got the wrong Bob?"

Gupte and Eliassi-Rad (2012) solve the inference problem of determining the weighted online implicit social network that gives rise to a set of observed events. They consider a set of users and a set of events where different users attend (possibly different) subsets of events with the possibility of several of these users simultaneously attending the same event. An example of an event represented in their study include those users who took at least one photograph at the same physical and temporal proximity of one another similar to that in *Color*. They then set out to determine how connected, which is represented by the strength of the tie in the network, any two users are based on a set of events. The only information they use in their approach is the knowledge that an event is attended by a known set of users. The underlying assumption in their approach is that there is an (implicit) relationship (e.g., interests) between any pairs of users who attend events which is indirectly based on an implicit social network.

Song et al. (2010) use online message threads to determine implicit social relationships among users who participated in those threads. They also introduce a visualization and interaction method that is suitable for exploring latent social relationships in message threads. They propose several algorithms and evaluate them using a Facebook dataset. Among the algorithms proposed and tested, the weighted harmonic rule mining with a root included sliding window showed the best performance. They claim that the visualization and interaction methods that they propose would enhance the usability of social network data in determining implicit social relationships. We develop a completely different model based on implicit social activities that is not based on the distance model as in the latent social network literature. Latent social network is based on a distance model that is tangentially related to the concepts discussed in this research. The reason for our inclusion of this reference is to prevent possible confusion between "latent" and "implicit" in the future.

Yang and Leskovec (2010) consider interactions among numerous participants and develop a Linear Influence Model. Rather than assuming the knowledge of a given social network and then modeling the diffusion by predicting which node will influence which other nodes in the network, they focus on modeling the global influence of a node on the rate of diffusion through the (implicit) network. They model the number of newly infected nodes as a function of other nodes that were previously infected. For each node, they estimate an influence function that quantifies the number of subsequent infections that can be attributed to the influence of that node over time. A nonparametric formulation of the model leads to a simple least squares problem that can be solved on large data sets. They then validate their proposed model on a data set comprising 500 million tweets and a data set comprising 170 million news articles and blog posts.

Frey et al. (2011) base their study on the premise that (a) explicit online social networks provide trusted social links and (b) implicit online social networks do not provide any trust guarantees while providing useful links. They then combine explicit and implicit online social networks to benefit from their complementary advantages - with the usefulness of implicit online social networks and the trust-worthiness of explicit online social networks. Their claim for the trustworthiness of links in an explicit online social network arises from these links connecting friends and co-workers, and other trusted parties. However, this may be questionable since links in an explicit online social network could be generated as a result of peer pressure, herd behavior, incentives, among others. Whereas their claim that implicit online social networks do not convey any kind of trust may be justified since the users on either side of an edge need not necessarily know each other in these networks. They extend their existing work on gossip overlays and propose Social Market, a solution to identify trusted social acquaintances. Their methodology, TAPS (Trust-Aware Peer Sampling), helps provide each user in an online social network with a set of neighbors who are simultaneously useful and trust-worthy.

Yoon and Zhou (2011) consider implicit online social network by adapting the social distance model and influence model to an implicit social network scenario. They then extend the basic model by incorporating the concept of multiple network paradigms.

3 Definition and Measurements

3.1 Measurement

Traditionally, the relationship between two individuals is characterized by their *direct* connections (with a 1 if it exists and a 0 when it doesn't exist). This is certainly the case when social relationships are measured with the most widely used consideration of directionality. The measure (using 0 or 1), however, only considers the *explicit* relationship. Essentially, 1 signifies that the two individuals are either friends or acquaintances through different types of roles in life. The value 0 signifies that the two individuals are not identified as having any direct connection. However, those two individuals can be *implicitly* connected through many channels. For example, one individual may be a loyal follower to another's blog, Facebook page, and twitter messages, without explicitly knowing the person. This individual may also be significantly influenced by these blogs, posts, opinions and product choices. We argue that in order to get a complete picture of the overall relationship between any two individuals' social networks, it is necessary to consider both the *explicit* and their *implicit* counterparts.

In general, social network relational directionality does not exist when the two parties on an edge in the network participate equally in a relationship - i.e., the relationship is mutual. While this may be true in some relationships, it is not always the case. This is somewhat different in traditional social network modeling where relationships are non-directional. We claim that in order to analyze the implicit flow of the network and its implication on business objectives, we need to fine-tune the measurement between any two network nodes. Clearly, there is a strong network imbalance in the social network, for example, between famous people (e.g., Barack Obama, Justin Timberlake) and their fans. It is possible for a person to claim social relationship with President Barack Obama simply through the fact that this person reads Mr. President's blog every day and through a Facebook book account-page link. This claim is, of course, not entirely correct. We argue that a better way is to model relational directions as well as strengths between any two nodes. Hence the premise of our modeling approach is to differentiate the direction and strength of connections in the network.

We consider and define *implicit* social relationship on the Internet, including an indicator from explicit social network (E) and another indicator from social activities among the individuals (A). E measures the conventional explicit connections, and A measures the implicit connections that are not considered in E.

Indicator from explicit social network E

Implicit social network model A→B→C $\Rrightarrow$ A→C

Indicator from social activities A

There exist several implicit social connections on the Internet, including:

1. Individual A observes individual B's activities
2. Individual A observes group B's activities
3. Both individuals A and B observe individual C
4. Both individuals A and B observe group C

The overall relationship between individuals i and j (X_{ij}) can be written as:

$$X_{ij} = f\{E[topology(i \leftrightharpoons j)], A_{i,j}\} \tag{1}$$

Here, f is a function of E and A, which can be in different formats incorporating different weights and even interconnected relationship between E and A. Using a linear function, it could be represented by the linear combination of both E and A, such as

$$X = \alpha + \beta_E X_E + \beta_A X_A \tag{2}$$

where $\beta_E X_E$ and $\beta_A X_A$ represent vector multiplication for explicit social network and social activities. $\beta_A X_A$ could be simply a vector to measure the degree to which i and j are connected, such as the number of years they have known each other, if they are friends on Facebook, if they are connected on Linkedin, etc. $\beta_E X_E$ is a vector that measures the degree of the four types of implicit connections discussed earlier. Construction of $\beta_E X_E$ may not be straightforward depending on the concentration of activities between the individual and groups. Both $\beta_E X_E$ and $\beta_A X_A$, nevertheless, can be empirically estimated based on observations of individuals' and groups' behavior and activities. Such a measurement structure allows us to generate better and deeper insights on online consumer behavior by considering the influence of both *explicit* and *implicit* social interactions. This structure also opens up the opportunity to take advantage of large-scale online (implicit + explicit) social network data analysis, which has thus far been limited to only explicit relationships.

The "observation" process itself is regulated by certain spatio-temporal sampling requirements. The goal of observation is to maintain a full picture of the implicit social activities without losing much relevant information. A simple example is to "observe" an event when an activity is recorded, such as a browsing record, a sales transaction, a reply, a post, etc. If we consider automated context-aware IoT/RFID tracking/tracing of continuously moving objects in a live environment, the sampling of spatio-temporal data follows Nyquist-Shannon's theorem. The difference between this type of implicit social relationship and an explicit social network is that the two entities don't have to register in the system (as in any existing Internet social network) to be explicitly related.

3.2 Social network matrix

When implicit connections are observed/measured among individuals in the community, the directional social relationship can be represented using x_{ij}, which represents the directional relationship from node i towards node j, for any $i \in [1, n]$.

$$x_i = [x_{i1}, x_{i2}, \cdots, x_{in}] \tag{3}$$

In social network matrix, we use the same concept of “node” to represent n individuals as in a traditional social network scenario. Unlike in the traditional scenario, the social network relationship between two nodes are no longer denoted as $\{0, 1\}$. Instead, the relationship is directional and x_{ij} and its strength are represented by individual points in $[0, 1]$.

3.2.1 Outbound relationship

Following the notation of directional social network relationship, for each node in the social network, we therefore compute the aggregate outward relationship of individual i as:

$$x_i = \sum_{j=1}^{n} x_{ij} \tag{4}$$

Each node’s outbound relationship can be explained as its overall interest towards all the nodes in this network, including itself. We then define the absolute outbound relationship $\hat{x_i}$ as.

$$\hat{x_i} = \sum_{j=1, j \neq i}^{n} x_{ij} \tag{5}$$

where $\hat{x_i}$ represents node i’s outbound relationship without including itself.

3.2.2 Inbound relationship

If we consider only the inward attention towards individual j, we represent this as:

$$Y_j = [x_{1j}, x_{2j}, \cdots, x_{nj}]^T \tag{6}$$

The inbound relationship of individual j can then be stated as:

$$y_j = \sum_{i=1}^{n} x_{ij} \tag{7}$$

Node j's inbound relationship y_j can be interpreted as the total interest that it receives from all nodes in the network, including j itself. We then define j's absolute inbound relationship as

$$\hat{y}_j = \sum_{i=1, i \neq j}^{n} x_{ij} \tag{8}$$

which represents the overall inbound relationship excluding the one from j itself.

3.2.3 Social Network Matrix

The complete social relationship can be represented by a symmetric matrix X_n such that

$$[X]_n = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1n} \\ x_{21} & x_{22} & \cdots & x_{2n} \\ \vdots & & \ddots & \\ x_{n1} & x_{n2} & \cdots & x_{nn} \end{bmatrix}$$

with each cell x_{ij} representing the directional relationship from node i towards node j.

Considering the possibility that x_{ij} can be zero for some edges in the directed graph, the social network matrix is asymmetric if $\forall j$, $x_{ij} = 0$, which essentially can be represented by subtracting the i^{th} row in the matrix. An asymmetric matrix signifies that in the community there are n individuals who are actively observing and m ($n \neq m$) individuals who are being observed.

$$[X]_{n,m} = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1n} \\ x_{21} & x_{22} & \cdots & x_{2n} \\ \vdots & & \ddots & \\ x_{m1} & x_{n2} & \cdots & x_{mn} \end{bmatrix}$$

Example and Implications

We now illustrate this concept through an example with three participants {A,B,C} in a cyber-social network. All three individuals know one another, and mutually and explicitly acknowledge this existing relationship. The explicit social network can be presented as an all-one matrix:

$$\begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix}$$

Given this social network topology, if social media wants to spread a piece of relevant information, the efficiency and effectiveness would be uniformly equal with any individual as the origin. Assume that after observing the implicit social activities for a period, we find that A always observes B and C but never posts any information, C actively posts information but never observes others, and B never observes anyone nor posts any information. The implicit SN matrix $[X]_{3,3}$ for {A,B,C} is:

$$[X] = \begin{bmatrix} 1 & 0.1 & 0.1 \\ 0.1 & 1 & 0.1 \\ 0.9 & 0.1 & 1 \end{bmatrix}$$

With this knowledge on the implicit social network relationship among the three individuals, the strategy for the social media will not be to equally choose among participants because of the obvious imbalance in implicit social activities among the players. In this case, the best strategy is to spread information starting from C, who will actively transfer this information to A. B, however, will be very difficult to reach in this social network without another channel or communication method.

3.2.4 Social Interest & Social Network Plate

Social activities are clustered by different social interests. For example, in many online communities that are based on unique interest, although they seem disparate from the outside they share several common features. Examples of this include photographers' discussion forum or an online computer DIY community. The Internet traffic as a common measurement indeed represents the accumulated social activities of many different social interests.

In order to provide a more accurate measurement and description of the Internet activities, we find it absolutely necessary to differentiate the overall set of social relationships based on their unique interests dimension, which we define as the *social network plate*.

We define a social network plate as the community with a common social interest (or social focus). Two plates k and l with different social interests are differentiated by a mapping function $f_{kl}(\cdot)$, such that the implicit social network relations from one plate can be shadowed upon another social network plate. In reality, we find examples of social network plate phenomena shadowing everywhere. For example, the social relationships and influence from an online photography community has a strong impact on another community that has an interest on photographic equipment. The impact becomes weaker towards the community of modern art, for example, although photography is a type of art presentation. However, the impact may be almost zero with a community on business school admissions.

Without loss of generality, consider the mapping function $f_{kl}(\cdot)$ as an angle $\alpha \in [0, 180]$ between any two different social network plates. The social network relationships from one social network plate (e.g., X_k) is therefore mapped to another (e.g., l) by $\overrightarrow{X_{kl}} = X_k \cdot cos(\alpha_{kl})$. The individual social relationship between any two nodes on a social network plate k is therefore strengthened by having social activities on another social network plate l, such that

$$
\begin{aligned}
x_{ij,k,l} &= x_{ij,k} + \overrightarrow{X_{lk}} && (9)\\
&= x_{ij,k} + x_{ij,l} \cdot cos(\alpha_{lk}) && (10)
\end{aligned}
$$

where i, j represent the two nodes and k represents the focal social network plate and l represents the supporting social network plate.

While social angle can be a convenient facilitator, we argue that in specific cases the practitioner should be able to find the most suitable function for $f_{kl}(\cdot)$, such as a linear function, logit, distance, etc. In the rest of this paper, we continue with the conceptualization of social network plate shadowing and the social angle without loss of generality. We define independency of two social network plates when they are orthogonal to each other, which signifies that the social relationship on one social network plate has no effect on the other as reflected by the effect between any pair of nodes.

Example and Implications

We continue with, and extend, “Example and Implications” from Section 3.2.3. We again consider the same social network with three participants {A,B,C}, but with two different social interests (e.g., architecture art and photography). We assume that all three players exist on both social network plates and explicitly acknowledge mutual relationships with one another on both plates.

The implicit social network matrix on one plate (architecture art) is observed and measured to be:

$$[X]_{art} = \begin{bmatrix} 1 & 0.1 & 0.1 \\ 0.1 & 1 & 0.1 \\ 0.9 & 0.1 & 1 \end{bmatrix}$$

The matrix on the photography social plate is:

$$[X]_{photo} = \begin{bmatrix} 1 & 0.3 & 0.5 \\ 0.2 & 1 & 0.3 \\ 0.1 & 0.8 & 1 \end{bmatrix}$$

If we assume that the shadowing function $f_{art \to photo} = x_{art} * 0.5$, the revised implicit social network matrix of photography interest becomes:

$$\widehat{[X]}_{photo} = \begin{bmatrix} 1 & .3 + .05 & .5 + .05 \\ .2 + .05 & 1 & .3 + .05 \\ .1 + .45 & .8 + .05 & 1 \end{bmatrix}$$

In this case, if a photo equipment retailer wants to make an effective social network advertisement to all these three individuals through multiple social platforms, it's better to use this revised matrix. In this example, we also want to emphasize that the distance (shadowing) function $f_{ij}(\cdot)$ normally is not reversible, which means that the impact of social interest A on social interest B shouldn't be considered the same (or reversible) as the impact from B on A. We thus introduce the social focus matrix in the next section to capture this set of relationships among various social interests.

3.2.5 Social Focus Matrix

A social focus matrix is defined to represent the relationship between any two social network plates, such that all the relationships among the different social network plates can be captured by matrix F:

$$F = \begin{bmatrix} f_{11}(\cdot) & f_{12}(\cdot) & \cdots & f_{1n}(\cdot) \\ f_{21}(\cdot) & f_{22}(\cdot) & \cdots & f_{2n}(\cdot) \\ \vdots & & \ddots & \\ f_{n1}(\cdot) & f_{n2}(\cdot) & \cdots & f_{nn}(\cdot) \end{bmatrix}$$

where function $f_{ij}(\cdot)$ represents a mapping of the social relationship from social network plate i to social network plate j. If we assume that the social interest mapping can be presented by the angle between any two social network plates, the original matrix can be modified as:

$$A = \begin{bmatrix} \alpha_{11} & \alpha_{12} & \cdots & \alpha_{1n} \\ \alpha_{21} & \alpha_{22} & \cdots & \alpha_{2n} \\ \vdots & & \ddots & \\ \alpha_{n1} & \alpha_{n2} & \cdots & \alpha_{nn} \end{bmatrix}$$

With a unique social interest, such as a business objective, we have a special social focus vector to represent the business objective: $A_\theta = [\alpha_{\theta 1}, \alpha_{\theta 2}, \cdots \alpha_{\theta n}]$, where $\alpha_{\theta i}$ is the angle from the focal social network plate i and the virtual business objective social network plate θ.

We observe that to determine the angle function between any two social network plates automatically from implicit social network activities would be very valuable if a uniform formula exists. In practice, this shadow function has to be discovered by defining specific models according to the unique setup in each problem. The directional mapping of each individual network pair is more context-specific rather than a simple general framework that can be applied without considering the exact situations involved. Therefore, in our model, we assume that we are able to discover the relationships between two social plates through expert knowledge, rather than a commonly applicable methodology that can be repeated automatically based on observation of implicit social activities.

Example and Implications

We continue with the three participants example from the previous sections to illustrate this concept. In order to obtain the most comprehensive social network matrix among the three players and to maximize the social media efficiency and effectiveness, it would make perfect sense for the business practitioners to obtain the "big data" from various social plates/platforms/interests. In order to effectively "observe" the social activities of all social network players, the data collection should have high velocity, variety and volume - the 3Vs that constitute the characteristics of modern conception of big data.

In order to achieve this goal, for a social network operator, one managerial implication is to manage multiple sites with different social interest or to promote inter-site collaboration among different service providers. We see great marketing potential from this practice to achieve more effective and efficient marketing communications with low cost and high return. We discuss some of the potential business benefits in the following sections.

3.3 Terminologies & Findings

Based on the definition of implicit social network matrix from the previous section, we now define three important parameters that can be used to measure the characteristics of social networks in its implicit form, which includes the absolute social relationship, interest-focused social relationship, and accumulated social centrality.

3.3.1 Absolute Social Relationship

Absolute social relationship (ASR) is measured by adding the absolute weights of relationship from all social network plates without considering the effects of shadow social interest from other social network plates.

$$ASR = \sum_{k=1}^{m} \sum_{i=1}^{n} \sum_{j=1}^{n} x_{ijk} \tag{11}$$

The value of ASR indicates the accumulated volume of both the static and dynamic social activities. However, it is not capable of providing differentiation information on social interests.

3.3.2 Interest-Focused Social Relationship

Interest-focused social relationship (IFSR) is measured by shadowing the social network matrix on the target social network plate θ.

$$IFSR_\theta = \sum_{i=1}^{m} X_i \cdot cos(\alpha_{i\theta}) \tag{12}$$

The target social plate θ can take a complete virtual form. The consumers' interest and actual demand of a specific product that appears in an online advertisement forms one type of virtual social network plate. $IFSR_\theta$ measures the total volume of the social activities that is related to the target social interest, conceived in the advertised product/service.

$$\widehat{IFSR}_\theta = \sum_{i=1, j \neq i}^{m} X_i \cdot cos(\alpha_{i\theta}) \tag{13}$$

Without considering the self-reflection from within the social network plate, $\widehat{IFSR}_\theta$ can be used to measure the the overall external social influence that is not directly associated with the target interest.

3.3.3 Accumulated Social Centrality

The accumulated social centrality (ASC) measures the global influence of a given social network plate.

$$ASC_i = \sum_{j=1}^{m} X_i \cdot cos(\alpha_{ij}) \tag{14}$$

ASC_i represents the centrality of a given social plate among communities of different social interests.

$$\widehat{ASC_i} = \sum_{j=1, j \neq i}^{m} X_i \cdot cos(\alpha_{ij}) \tag{15}$$

$\widehat{ASC_i}$ is the outward influence of this considered social network plate.

4 The Social Network Indicators

Traditionally, as seen in a majority of extant social network literature, the relationship between two entities are defined using the format $\{0, 1\}$, and are roughly defined without specification of its degree and its social interest. We argue that it's critical not only to specify the implicit degree of such a relationship, but also to specify the unique interest of such a relationship. For example, Alice and Bob may have a strong relationship based on research collaboration but almost zero relationship based on political interest. It signifies that if Mrs. Clinton wants to disseminate her next presidential election announcement through Alice to Bob, it's close to impossible because they may never discuss political topics. As a result, it's important to carefully differentiate the concept of relationship based on common social interest, such as the relationship between A and B through their shared interest on the art of photography, the relationship between A and B through their shared interest on photographic equipment technology or relationship between A and B through their shared interest in religion.

We therefore define an implicit social network of a special social interest based on accumulated shadowed relationships between two entities in the population (Figure 1). Let A and B represent the two entities and l represent this special social interest, x_{AB_l} represent the directional implicit social

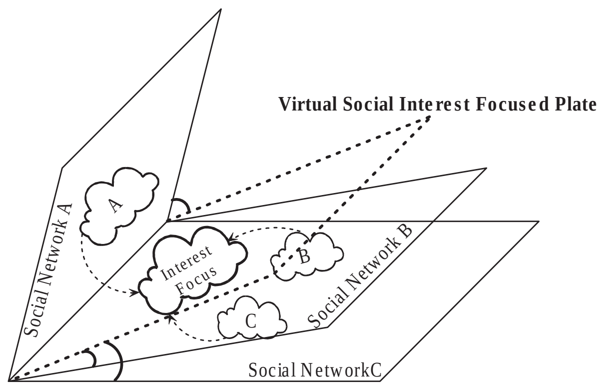

Figure 1: Implicit social network shadowing on a virtual social interest focused social network plate

relationship between A and B on social interest l, and

$$x_{AB_l} = \sum_{i=1}^{m} x_{AB_i} cos(\alpha_{il}) \tag{16}$$

Consequently, the implicit social network matrix of social interest l is:

$$[X]_l = \begin{bmatrix} x_{11_l} & x_{12_l} & \cdots & x_{1n_l} \\ x_{21_l} & x_{22_l} & \cdots & x_{2n_l} \\ \vdots & & \ddots & \\ x_{n1_l} & x_{n2_l} & \cdots & x_{nn_l} \end{bmatrix}$$

Based on $[X]_l$ we can draw the social network indicators using traditional network flow topology analysis.

To facilitate the analysis of implicit social network, we define an inverse relationship $d_{ij} = \frac{1}{x_{ij}}$, where d_{ij} represents the distance between node i and node j. Similarly, $[D]_l$ is represented as:

$$[D]_l = \begin{bmatrix} d_{11_l} & d_{12_l} & \cdots & d_{1n_l} \\ d_{21_l} & d_{22_l} & \cdots & d_{2n_l} \\ \vdots & & \ddots & \\ d_{n1_l} & d_{n2_l} & \cdots & d_{nn_l} \end{bmatrix}$$

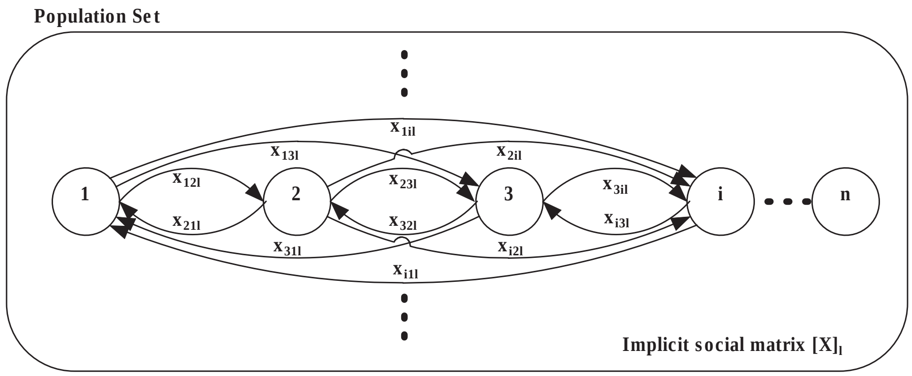

Figure 2: Interest focused Implicit Social Network

We now discuss three common dynamics that are of interest in implicit social networks: lowest cost (shortest path) problem, Max flow, bottleneck (Max flow-Min cut) theorem, knapsack problem.

4.1 Lowest Cost Problem

The edge relationship in an implicit social matrix represents the directional influence between any two entities in the network. The inverse of such a relationship, to the contrary, would represent the impedance or the cost of such influence. A firm that is interested in dissemination of information through such a network certainly has an incentive to minimize the cost to transmit information from a start node to the target node. Assuming that there is an operational cost c_{ij} on each edge, such that after a certain cost (e.g., $d_{ij} \cdot c_{ij}$) the firm interested in disseminating information (e.g., advertisement for a product/service) is able to transmit a piece of information from node i to node j. We formulate this problem based on an interest-focused implicit social network that accumulates implicit social network shadowing information from various social network platforms (plates). We model this as a directed graph $G = (N, A)$ with an edge cost c_{ij} associated with each edge d_{ij}

$$Minimize \sum_{(i,j)\in A} d_{ij}c_{ij} \tag{17}$$

subject to

$$\sum_{\{j:(i,j)\in A\}} d_{ij} - \sum_{\{j:(i,j)\in A\}} d_{ji} = \begin{cases} n-1 & i = \{s\} \\ -1 & \forall i \in N - \{s\} \end{cases} \tag{18}$$

$$d_{ij} \leq 0 \;\; \forall (i,j) \in A \tag{19}$$

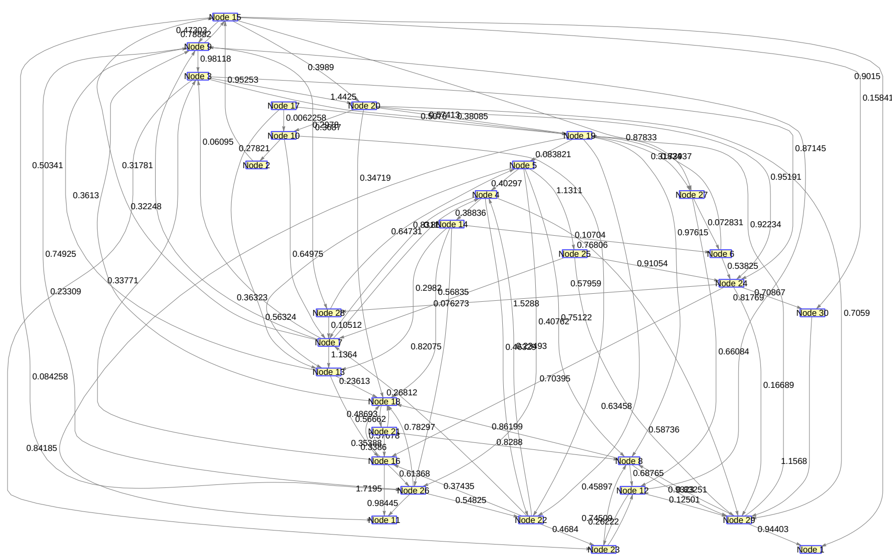

Figure 3: An interest focused implicit social network with communication cost matrix

Figure 3 illustrates an implicit social network of 30 nodes and 100 non-zero relationships. Each edge is characterized by a communication cost factor to ensure successful directional communication. Figure 4 presents the low cost dynamics of the social network presented in Figure 3.

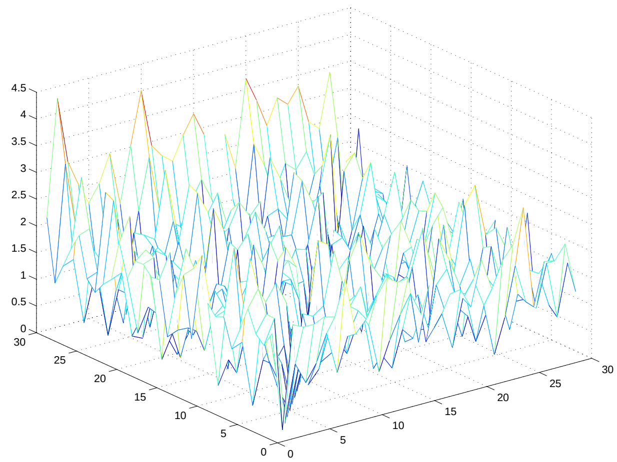

Figure 4: Plot of the lowest cost matrix

We observe two variations of this problem: 1. find the lowest cost from one node to all other nodes and 2. find the lowest cost from every node to every other node. Assuming the source node comes from a business entity, the first problem could provide an indicator for the business practitioners to evaluate the efficiency of their practice. The second problem can be considered simply as a general indicator of an individual social network or a set of social networks.

4.2 Maximum Flow Problem

It is not uncommon to assume that the Internet has the capacity to handle an unlimited amount of data and information. Clearly, this is not true since the information flow on the Internet has a capacity constraint. In other words, in a given time period, not all data/information can be transmitted to the target node although they can be covered without the time window.

From a business practitioner's perspective, it's important to realize this constraint on the Internet when there is a need to disseminate a large amount of information in a short time period. Consequently, the maximum "capacity" of the social network platform needs to be considered.

Considering that $[X]$ is measured mostly based on the implicit activities in the social network, $[X]$ represents the directional capacities of information flow between any pair of nodes. For instance, x_{ij} measures the implicit relationship from node i to node j that also represents the capacity of directional communication from node i to j. For disseminating information with the goal to maximize the flow of information in the social network, the problem would be to decide how to send information between the source node and a sink node in a capacitated network and its maximum capacity. The source node could represent the practitioner who desires to disseminate information and the sink node could represent the targeted audience. The optimization problem can be written as:

$$Maximize\ v \tag{20}$$

subject to

$$\sum_{\{j:(i,j)\in A\}} x_{ij} - \sum_{\{j:(i,j)\in A\}} x_{ji} = \begin{cases} v & i = \{s\} \\ 0 & \forall i \in N - \{s\ and\ t\} \\ -v & i = t \end{cases} \tag{21}$$

$$0 \leq x_{ij} \leq u_{ij} \quad \forall (i,j) \in A \tag{22}$$

There exist several algorithms to solve the above-proposed problems, such as the augmenting path algorithms and the preflow-push algorithms. We also argue that in a social network, the maximum value of the information flow from a source node to a sink node equals the minimum capacity (a.k.a. the bottleneck) among all source-sink cuts.

4.3 Budgeting Problem

Social networks have capacity and cost limits for information transfer. Consequently, practitioners must prioritize their goals with respect to communication of information in the network in order to select an effective combination of information/tasks according to cost and capacity constraints.

For cost constraints, it is necessary to decide which pieces of information to be disseminated should be included in the next project. The choice is among g pieces of information, where each piece of information i has a dissemination cost c_i (or size w_i) and a utility u_i. The objective is to maximize the utility of the information disseminator's project subject to the cost or size limitation that she can transmit no more than C (or in size, W).

$$Maximize \sum_{i=1}^{g} u_i x_i \tag{23}$$

subject to the cost

$$\sum_{i=1}^{g} c_i x_i \leq W \tag{24}$$

$$x_i = \{0,1\} \quad \forall i \tag{25}$$

or subject to the capacity

$$\sum_{i=1}^{g} w_i x_i \leq W \tag{26}$$

$$x_i = \{0,1\} \quad \forall i \tag{27}$$

Equations 23-27 represent a basic setup for this budgeting problem. This can be easily modified later as necessary and appropriate to anchor a specific social network flow problem.

5 Discussion

We proposed a new framework to evaluate the characteristics of a set of social networks. In this framework, a social network matrix is proposed to measure the implicit relationships among the entities in various social networks. Based on this, we derive several indicators to characterize the dynamics of explicit social networks along with their implicit counterparts. We incorporated implicit social network in a traditional network flow context to develop the key network performance indicators such as lowest communication cost, maximum information flow, and budget constraints.

Based on network analysis and proposed implicit social network framework, we are now able to analyze and discover not only the topology that is different from a traditional explicit $\{0, 1\}$ social network but also the key nodes, the bottleneck, and optimization strategy with respect to balancing cost and capacity for the social network. For future research, it might be interesting to identify the propagation pattern of information flow in social networks. Many problems associated with social network management can be approached using the proposed social network framework that considers both implicit and explicit social networks as well as a set of such networks across several social network plates. For example, we can differentiate the broadcasting/consumer targeted strategies for Internet advertising, simply by varying the cost matrix in the lowest cost problem. The model can also be used to solve information propagation problems.

The implicit social network measurement and matrix can be extended to answer many existing questions in current social network research. Based on the proposed setup, a possible future research can be directed to extend this research to the classic Hotelling model in the social network domain, by assuming a social network plate with a unique social interest (or a certain marketing interest) "X_θ". While there might physically exist a social community that matches exactly with θ, it can also represent a virtual social network plate. By modifying the classic linear Hotelling model, we can define: (1) two social network plates 1 & 2; (2) a sponsor with a unique business interest θ that can be implemented on either social network plate 1 or 2; (3) the social angles from θ to social network plates 1 and 2 to be respectively θ_1 and θ_2, and $\theta_1 + \theta_2 = \Theta$. We leave this as an exercise for future research.

References

[1] Aedo, I., Diaz, P., Carroll, J. M., Convertino, G., & Rosson, M. B. 2010. "End-user oriented strategies to facilitate multi-organizational adoption of emergency management information systems." *Information Processing & Management*, 46: 11-21.

[2] Frey, D., A. Jegou, A.-M. Kermarrec. 2011. Social Market: Combining Explicit and Implicit Social Networks. *International Symposium on Stabilization, Safety, and Security of Distributed Systems.*

[3] Gupte, M., and T. Eliassi-Rad. 2012. Measuring Tie Strength in Implicit Social Networks. *Proceedings of the 3rd Annual ACM Web Science Conference (WebSci)*. 109-118.

[4] Lattanzi, S., and Sivakumar, D. (2009, May). Affiliation networks. In Proceedings of the 41st annual ACM symposium on Theory of computing (pp. 427-434). ACM.

[5] Nauerz, A., and G. Groh. 2008. Implicit Social Network Construction and Expert User Determination in Web Portals. *Proceedings of the AAAI Spring Symposium on Social Information Processing*, 60-65.

[6] Roth, M., T. Barenholz, A. Ben-David, D. Deutscher, G. Flysher, A. Hassidim, I. Horn, A. Leichtberg, N. Leiser, Y. Matias, R. Merom. 2011. Suggesting (More) Friends Using the Implicit Social Graph," *International Conference on Machine Learning (ICML)*.

[7] Smith, M., C. Giraud-Carrier and N. Purser. 2009. Implicit affinity networks and social capital. *Information Technology and Management*, 10(2-3), 123-134.

[8] Song, M., W. Lee, J. Kim. 2010. Extraction and Visualization of Implicit Social Relations on Social Networking Services. *Proceedings of the Twenty-Fourth AAAI Conference on Artificial Intelligence (AAAI-10)*, 1425-1430.

[9] Wasserman, S. and K. Faust. 1994. Social network analysis: Methods and applications. Cambridge university press.

[10] Yang, J. and J. Leskovec. 2010. Modeling Information Diffusion in Implicit Networks. *Proceedings of the IEEE 10th International Conference on Data Mining (ICDM)*, 599-608.

[11] Yoon, E.J. and W. Zhou. 2011. Mining Implicit Social Network with Context-Aware Technologies. *Workshop on eBusiness (WeB)*. LNBIP 108, 3-8.