Tie-Strength and Strategies in Social Capital Management

Sungpack Hong
Pervasive Parallelism Laboratory
Stanford University
hongsup@stanford.edu

Keywords Social Network, Tie Strength, Job Information, Social Capital Management

1. Introduction

One of most influential sociology studies is the concept of strength in interpersonal relationship and its role in information spread-out through social networks. In sociology, tie-strength is a mathematical model that represents the degree of intimacy between two people; strong ties, as for a rough classification, correspond to close friendship and weak ties to acquaintances.

The surprising discovery, however, was that critical information, such as job position, tends to flow through weak-ties rather than strong-ties. Once after Granovetter [8] first reported this phenomenon in his Ph.D work through extensive empirical studies on newly hired people, it drew great attention from many researchers in sociology and economics who tried to find a natural explanation for it.

There is a group of convincing explanations [7, 9, 10] which relate weak-ties with the structure of underlying social network. In these arguments, strong-ties have structural meaning such that two people (nodes) in the social network are highly likely to have a friend (another node) in common, if there is a strong-tie between the original two people. More accurately, if there is a strong tie between node A and node B, and another strong tie between node A and node C, then it is highly likely that there exits a tie, at least weak one, between node B and node C (Strong Triadic Property). The mathematical consequence of this explanation is that weak ties serves as the edges which bridge different communities or cliques in the network together; novel information naturally comes from such foreign links.

In contrast, there are another set of arguments from economists [2, 4, 11, 12] who tried to set up probabilistic models that explain the behavior of job market with respect to the tie-strength phenomenon. These studies put more focus on the behavioral aspect of tie-strength. As an example, in the original work from Boorman [2] each individual has finite amount of resource to keep all the interpersonal ties while a strong-tie costs more than a weak-tie. Thus, weak-ties make it possible for each node to keep many number of ties, which in turn increases the chance to receive the information. However, it is also assumed that the probability of information flow from a node to another is proportional to the tie-strength between these two nodes. Thus Boorman's model [2] claims that in the absence of competitions a node can gain much information through its many weak ties, while the usefulness of weak-ties fades out under high degree of competitions since they are not strong enough to win the information over other nodes.

However, the above approaches by the economists typically assume that the underlying social network graph is a regular and infinite one; based on such assumption, it is possible to conduct mathematical analysis using their model in order to predict the properties of the stable state, or equilibrium. Although this approach, favored in economics, provides a very useful insight and powerful mathematical analysis tool to understand the nature of complex real-world phenomenon, such an assumption of regular infinite network is mostly unrealistic. To the quite contrary, it has been shown that real-world social networks have very small diameters [13] and very skewed degree distribution [1].

Therefore, it is an interesting question to ask which of two aspects of tie-strength, structural and behavioral, have higher influence on the information transfer in real social network instances, if any. In this study, we first aim to answer this question. Our approach is similar to the previous approaches in socio-economics [2, 4, 11, 12], since we start from a reasonable probabilistic model that captures behavioral aspects of interpersonal tie-strength, and use its consequence to explain the nature of real-world phenomenon. However, unlikely to those previous approaches, instead of doing mathematical analysis, we conduct a direct Monte-Carlo simulation using the model on realistic social network instances in order to accommodate the structural effect of the networks.

Another interesting question we want to address in this study is what would be a good strategy for each individual to maximize information reception through the social network. One basic question regarding Boorman's claim [2] is whether to make many weak ties or to make small number of strong ties. More interestingly, if the number of one's neighbors is somehow pre-determined, how one should allocate one's resource among those neighbors: either to put more resources onto high-degree neighbors or onto low-degree neighbors. We also answer this question through the same Monte-Carlo simulation.

This work makes the following contributions:

- This study (dis-)verifies the previous hypotheses on the relationship between tie-strength and the information flow, using direct simulation on real-world network instances, which is to the best of our knowledge, the first of such an approach.
- We propose a fast Monte-Carlo simulation method which enables to conduct such a simulation on a large network instances in a relatively short amount of time.

In following sections, we first explain our models for tie-strength and information flow (Section 2) and discusses a method for fast Monte-Carlo simulation (Section 3). Each specific hypothesis can be found before we present experimental results for it (Section 4). We also provide discussions for further extension of this work (Section 5) before we concludes.

2. Model for tie-strength and information flow in social networks

In this study, we use a probabilistic model for tie-strength and information flow on social networks that closely resembles a similar model of Calvo-Armengol and Jaction [3].

In our model, each node in the network has a unit amount of resource to keep the interpersonal ties with its neighbor. Our approach assumes that the existence of connection between two nodes is determined a priori, while each node has flexibility to allocate its resource arbitrarily among its neighbors. The tie-strength from node i to node j is proportional to the resource invested from node j to node i. More specifically,

$$\tau_{ij} = \frac{R_{ji}}{\sum_{j' \in nbr(i)} R_{j'i}} \tag{1}$$

Here, τ_{ij} is the tie strength from i to j and R_{ji} is the resource and R_{ji} is the resource spent from node j to node i. $R_{ji} > 0$ if there exists a tie between j and i, $R_{ij} = 0$ otherwise. Also, for each node j, $\sum_{i \in nbr(j)} R_{ji} = 1.0$.

Note that in this model, tie-strengths are asymmetric between two nodes: $\tau_{ij} \neq \tau_{ji}$. This is the natural consequence of asymmetry of underlying social network instances. For example, if a million-degree node is connected to a million single-degree nodes, it is not likely for all of those tie to be strong, even though each single-degree node puts all of its resource to its only relationship. In a behavioral interpretation, this corresponds the situation where emotional intensity between two people is not the same reciprocally, which can be observed frequently in everyday experiences [5].

We model the influence of tie-strength on information flow within a social network with following steps:

1. Each node generates unit amount of information with probability of α.
2. Each node needs information with probability of β. Say $i \in B$, if node i needs the information.
3. When a node i have the information but does not need it, it firsts looks at the neighboring nodes that require the information. If there are such nodes, the information is propagated one of such nodes (say node j) with probability in equation (2).

$$\frac{\tau_{ij}}{\sum_{j' \in nbr(i) \cap B} \tau_{ij'}} \tag{2}$$

4. If there is no such neighbor, node i now chooses K distinct nodes among its neighbor randomly again with probability in (2) and gives the information to those nodes. However, the transfered information now has value 1/K times the original value. (This step models 'ask for asking' in the real world.)
5. When a node receives the information while it does not need the info, with probability of γ the node repeats step 3 – 5. Otherwise the information is simply lost.

3. A Method for Fast Monte-Carlo Simulation

The effect of information flow model in the previous section on large real-world network instances. is hard to be analyzed mathematically. Instead, we discuss a method to do Monte-Carlo simulation of the model on such networks in this section.

There is a very intuitive way to perform a naive Monte Carlo(MC) simulation; at each time step, one chooses α fraction of nodes that generate the information and β fraction of nodes that need the info, and simulate information flows of these nodes. this process is repeated until the simulation covers sufficiently large number of situations.

However, such a naive MC simulation requires exponential number of time steps to reasonably approximate the result of the probabilistic model on a large sized network instances. This is because every different combination of choices for information source and sink produces different result; one needs many simulations that covers all the possible combinations, multiple times. Note that Calvo-Armegol and Jackson performed this naive MC simulation with a model similar to ours [4], while their simulated network size was very small (~ 20).

Instead, we *compute* the average amount of information that is received by each node using flows and conditional probability. More specifically, the quantity we want to compute is expressed as follows:

$$F(i) = E[\text{flow goes into i} \mid \text{ i needs info}]$$

, where E[] denotes the expected value.

However F(i) can be re-written as:

$$\begin{aligned} F(i) &= \sum_{j \in nbr(i)} \sigma(j,i) * X(j) \\ \sigma(j,i) &= \text{prob}\{\text{info from j to i} \mid \text{i needs info}\} \\ X(j) &= E[\text{excess flow from node j} \mid \text{ j needs not info}] \end{aligned}$$

In other words, F is the summation over all neighbors of the product of probability of information flow(σ) from that neighbor and the amount of excess information(flow) from the node.

Finally, X(j) can also be formulated as following:

$$\begin{aligned} X(j) &= \alpha + Y(j) * \gamma \\ Y(j) &= E[\text{flow into node j} \mid \text{node j needs not info}] \\ &= \sum_{w \in nbr(j)} \delta(w,j) * (1-\beta)^{deg(w)} * X(w)/K \\ \delta(w,j) &= \text{prob}\{\text{info from w to j} \mid \text{no nbr of w needs info}\} \end{aligned}$$

In other words, the available excess flow from node j is the sum of the amount of generated information and expected flow going into node j when j needs not information multiplied by gamma. The latter can be again computed by summation over all neighbors of the product of probability of information flow(δ) and the amount of excess information reduced by K.

We compute the probability σ and $delta$ per node Monte-Carlo simulation with following algorithm: (Skipped in this version of the report due to time limit)

Finally, X can be calculated using an iterative method, which enables to compute F as well. Algorithm in Figure 1 shows our whole computation steps. Note that we only need independent MC simulations for O(M + N) times.

4. Hypotheses and Experiemtal Results

In this section, we perform the simulation of our information flow model in previous sections with real social network instances and present their results. We used two real network instances from a public social network repository [6]:ca-GrQc (collaboration network of Arvix General Relativity) and p2p-Gnutella05 (gnu peer to peeer network from August 5 2002) [1].

4.1 Uniform Resource Allocation

In first set of simulation, we assume uniform resource allocation; that is, each node allocates the same amount of its resource to every neighbor. As for parameter values as in the Table 1.

[1] The result of my third dataset, wiki-Vote (Wikipedia who-votes-on-whom), was corrupted and could not be re-generated within the given time frame

```
G: Graph instance

// compute sigma and delta
foreach node n of G {
  foreach node m of n.nbrs {
     compute_sigma_with_MC(n, m);
  }
  compute_delta_with_MC(n, n.nbrs);
}

// compute X
foreach node n of G {
   X[n] = alpha;
}
bool finished = false;
while (!finished) {
  finished = true;
  foreach node n of G {
     float sum = 0;
     foreach node m of n.nbrs
        sum += delta(m,n) *
             (1-beta)^(m.deg) * X[m]/K * gamma;
      X_next[n] = alpha + sum;
      if (X_next[n] - X[n] > epsilon)
         finished = false;
   }
   foreach node n of G
      X[n] = X_next[n];
}

// Compute F
foreach node n of G {
    float sum = 0;
    foreach node m of n.nbrs
       sum += sigma(m,n) * X(m);
    F(n) = sum;
}
```

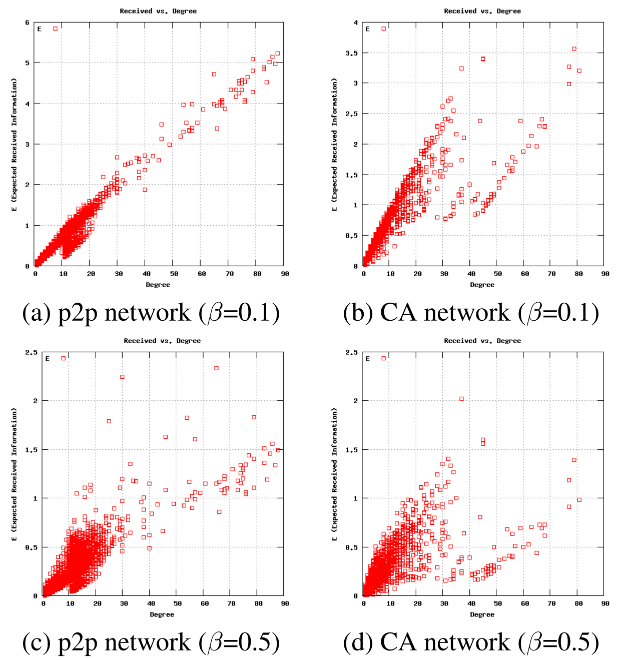
Figure 1. Our Monte Carl simulation method.

parameter name	value
α	0.1
β	varied (0.5 and 0.1)
γ	0.1
K	3

Table 1. Parameter Values in Simulation

With these experimental setup, we now observe the relationship between tie-strength and the information flow from the simulation result. First, we verify the claims about the behavioral aspects of tie-strength and information flows.

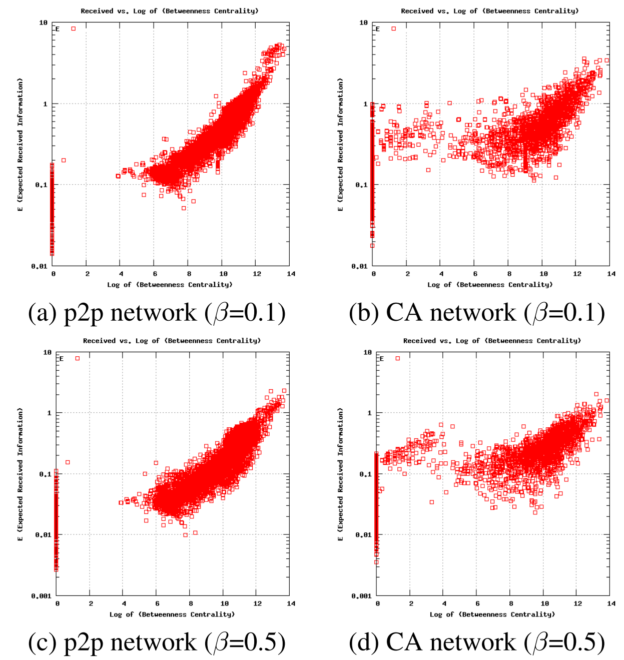
- **Hypothesis No.1 (Boorman [2])**: There is a positive correlation between the degree of a node and the amount of information received by the node when the degree of competition (β) is small.
- **Hypothesis No.2 (Boorman [2])**: There is a negative or no correlation between the degree of a node and the amount of information received by the node when the degree of competition (β) is high.



(a) p2p network (β=0.1) (b) CA network (β=0.1)

(c) p2p network (β=0.5) (d) CA network (β=0.5)

Figure 2. Verification of Hypothesis No.1 and No.2: X-axis is the degree of node and Y-axis is the amount of information received by the node.



(a) p2p network (β=0.1) (b) CA network (β=0.1)

(c) p2p network (β=0.5) (d) CA network (β=0.5)

Figure 3. Verification of Hypothesis No.3: X-axis is log of (betweenness centrality +1) of the node and Y-axis is the amount of information received by the node.

The first hypothesis claims that the weak-ties in a high-degree node helps to receive more information due to their plurality since there are little competitions. Figure 2.(a) and (b) verifies this hypothesis. In both graph instances, we see strong correlation (almost linear) between degree and received information.

The second hypothesis claimed that under high contention the usefulness of weak-ties fades away since they do not win the information through competition. However, our simulation result shown in Figure 2.(c) and (d) did not agree with this claim. In the figure, compared to Figure 2.(a) and (b), although the correlation has blurred somewhat, the figures still showed a certain degree of correlation between node degree and information received, even under high contention value (β=0.5).

The above disagreement can be explained in following way. Although each weak ties from a high-degree node receives little information from the connected node, the total sum of such little information builds up a substantial value. Also, by the nature of underlying social network, high-degree nodes tend to be connected with low-degree nodes. Since there is little competition around such low-degree nodes, there is still a good chance of receiving information from those nodes even with weak ties.

Now we look at the claims about the structural aspects of weak-ties and information flow.

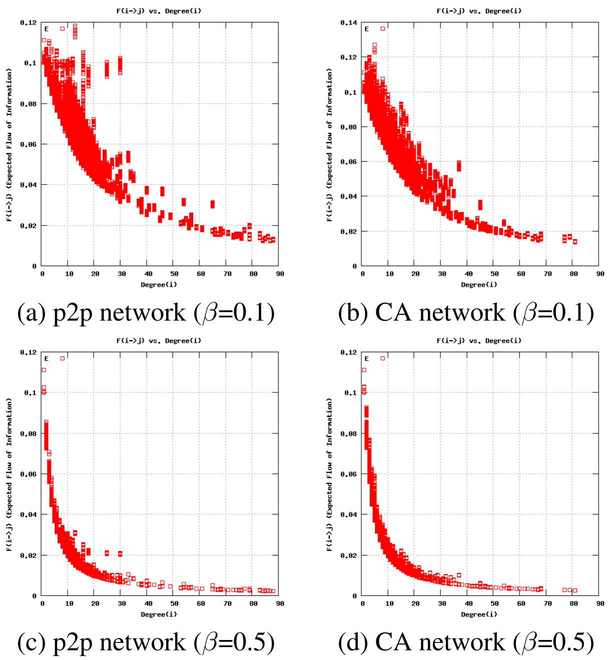
- **Hypothesis No.3 (Granovetter [9])**: There is a positive correlation between the betweenness centrality of a node and the amount of information received by the node (regardless of the degree of competition (β))

The third hypothesis emphasis on the effect of weak-tie as bridging edges, or contribution to increasing betweenness centrality(BC). Our simulation results in Figure 3 again confirms this hypothesis. In the figure, we plotted log of BC against log of information received, which again showed linear relationship. This implies BC and amount of information have power-law relationship, while coefficient differs depending on network structure. Also this relationship is relatively insensitive to the degree of competition.

Now we make further observations on how information flows in the networks.

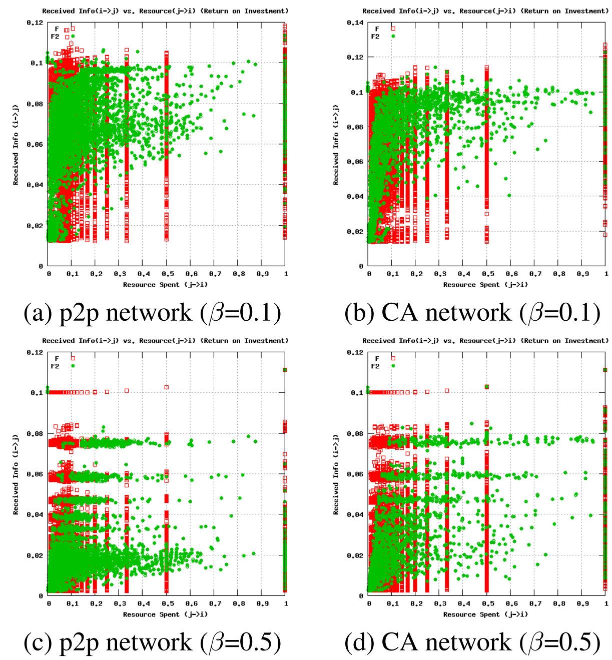
- **Hypothesis No.4** : There is a positive correlation between the degree of a node and the amount of information received from the node

The fourth hypothesis takes a look at information comes more from which kinds of nodes. The intuition was that since a high-degree node receives more information as seen in Figure 2, it would provide more information to its neighbors. The simulation result in Figure 4 was, however, the opposite of this intuition. This is because due to the competition around the high-degree nodes, the information comes out of it is small. Bluntly put, famous people are too busy for you. We can notice that the amount of information from high-degree nodes becomes even smaller when competition(β) is high.



(a) p2p network (β=0.1) (b) CA network (β=0.1)
(c) p2p network (β=0.5) (d) CA network (β=0.5)

Figure 4. Verification of Hypothesis No.4: X-axis is degree of a node and Y-axis is the amount of information comes out from the node through a link.



(a) p2p network (β=0.1) (b) CA network (β=0.1)
(c) p2p network (β=0.5) (d) CA network (β=0.5)

Figure 5. Verification of Hypothesis No.5: X-axis is the amount of resource put into a node, and Y-axis is the amount of information comes from the node

4.2 Non-Uniform Resource Allocation

Now, we change our attention to a slightly different issue of resource allocation. Although we assume that the network structure (i.e. degree and centrality value of each node) is exogenous or pre-determined, each node still has the flexibility to allocate its resource differently among neighbors.

Thus, the interesting question is what would be a good (or the best) strategy to maximize the received information. Should a node put more resources on high-degree nodes or on low-degree nodes? To answer this question, we devised a few strategies for allocations and simulated their results. Specifically we randomly choose ϕ fraction of node to apply the strategy and measure information received from each neighbor versus the amount of resources spent to the neighbor.

- **Hypothesis No.5** : There exists a resource allocation strategy that induces better positive correlation between information received and resource spent.

We present the result of one of our strategies that gave the best result. The strategy is named as *Blue Ocean*, which is as following rules:

1. Allocate minimum (10^{-8} in our simulation) resource to single-degree neighbors.
2. Allocate remaining resources to other neighbors inversely proportional to their degree.

In other words, this strategy weights more on low-degreed nodes to increase chances to win competitions.

The experimental result for this hypothesis and strategy can be found in Figure 5. In the figure, the red box is the investment-return plot for the uniform resource allocation scheme, which shows no correlation. On the other hand, the green dots are resulting plots for Blue Ocean strategy. As can be seen in the plots, the strategy induces better ROI than uniform allocation; the ROI correlation drops under high contention, since there might happen cases that the node does not win the competition with the enhanced bidding of this strategy. However, overall this strategy is far better than uniform allocation.

5. Future Works

Although this work presented a novel approach and gives many insights on the problem, we were not able to finish to explore all the ideas.

The experiments can be improved to include more network instances. However, since the results showed quite similar shapes from two very different instances, we believe our result will hold for wide range of networks.

We have not explored the effect of other parameters like γ, K.

Writing was done in haste and do not include all the organized analysis of the results.

6. Conclusion

In this study, we performed a direct Monte-Carlo simulation on a large real-world social network instances, to explore the relationship between tie-strength and information flow in those networks. We devised a fast MC simulation method to expedite such an approach

Our results first confirmed that both the behavioral aspects and structural aspects can be observed in real-world social network instances. However, unlikely to a previous model, our simulation showed that high-degree nodes receive more information even under high contention, since most of data comes from low-degree nodes. Finally, we showed that there exists a better resource allocation strategy that results in better data reception per resource investment than uniform allocation.

Acknowledgments

I appreciate Rae Kim, processor of Media and Communication Department at University of Dayton, for his valuable discussion with me on this topic.

References

[1] A. Barabási and R. Albert. Emergence of scaling in random networks. *Science*, 286(5439):509, 1999.

[2] S. Boorman. A combinatiorial optimization model for transmission of job information through contact networks. *The bell journal of economics*, 6(1):216–249, 1975.

[3] A. Calvo-Armengol and M. Jackson. Social networks in determining employment: Patterns, dynamics, and inequality. *American Economic Review*, 94(3):426–54, 2004.

[4] A. Calvo-Armengol and M. Jackson. The effects of social networks on employment and inequality. *The American Economic Review*, 94(3):426–454, 2004.

[5] J. Cham. Average time spent composing one e-mail. http://www.phdcomics.com/comics/archive.php?comicid=1047/.

[6] J. Cham. Average time spent composing one e-mail. http://www.phdcomics.com/comics/archive.php?comicid=1047/.

[7] D. Easley and J. Kleinberg. *Networks, Crowds, and Markets: Reasoning about a Highly Connected World.* Cambridge University Press, 2010.

[8] M. Granovetter. The strength of weak ties. *American journal of sociology*, 78(6):1360–1380, 1973.

[9] M. Granovetter. The strength of weak ties: A network theory revisited. *Sociological theory*, 1(1):201–233, 1983.

[10] M. Granovetter. The impact of social structures on economic development. *Journal of Economic Perspectives*, 19(1):33–50, 2004.

[11] J. Montgomery. Social networks and labor-market outcomes: Toward an economic analysis. *The American economic review*, 81(5):1408–1418, 1991.

[12] J. Montgomery. Weak ties, employment, and inequality: An equilibrium analysis. *American Journal of Sociology*, 99(5):1212–1236, 1994.

[13] D. Watts and S. Strogatz. collective dynamics of 'small-world' networks. *Nature*, 393(6684):440–442, 1998.