Suppression of Information Diffusion in Social Network Using Centrality based on Dynamic Process

Eiichi Takazawa and Norihiko Shinomiya

Graduate School of Engineering
Soka University
1-236 Tangi, Hachioji, Tokyo, Japan
Email: shinomi@soka.ac.jp

***Abstract*—Individual activities propagate on social networks and had a large impact on our society. For example, incitement acts such as hoaxes, widely propagated through social media, gave unnecessary confusion and uneasiness to many people. The purpose of this study is to propose an edge centrality index in a network considering the propagation of activities through analysis. Our previous studies have proposed an evaluation method that quantifies the edge importance based on an activity propagation model. The model represents the propagation by an equalization process of the variable amount given to each vertex. This paper experimentally shows that the information diffusion can be suppressed by using the edge importance measure. The experiment verifies that the range of information diffusion becomes smaller than that before deleting some edges from the network based on the importance measure.**

***Keywords*–Centrality; Graph Theory; Information Diffusion.**

I. Introduction

The development and dissemination of information network technologies have accelerated online communications among individuals and dispatch of information at the individual level. In particular, individual activities also have propagated throughout the world due to the advancement in social media. As a result, the impact of individual activities on society has been increasing. For instance, in the so-called Arab Spring, democratization movement spread through social media. That led to huge demonstration activities that occurred in many countries [1]. In the future, it is expected that technological innovation such as blockchain further facilitates interaction among individuals, complicates mutual influence, and makes the scale spread out. Therefore, it is necessary to analyze not only a network itself but also the behaviors of the people connecting to the network.

On the other hand, a variety of *centralities* have been theoretically and experimentally studied in the field of network analysis and graph theory [2]. Each centrality quantifies how important a vertex or an edge is in a graph that abstracts the corresponding network. There are different measures of centrality depending on how to define the criteria on importance. For example, degree centrality [2] is based on the number of edges connected to vertices. As the significance index of centrality, it is assumed that vertices with higher degree strongly contribute to spreading information. Betweenness centrality [3] is defined as a ratio of an edge existing on the shortest path between arbitrary two vertices. It is assumed that information tends to be transmitted along the shortest path. These conventional centralities have been applied in a wide range of fields such as human relationship analysis and information communication network design. However, these measures are defined only based on the *static structure* of the network such as the degree and the shortest path, thereby not suitable for analyzing a *dynamic process* such as activity propagation.

Therefore, the purpose of this study was to propose a new edge centrality index in a network considering the propagation of activities through analysis. The proposed centrality is expected to be used as a clue of prevention on the diffusion of computer viruses, false propaganda, and online flaming, since it can measure the edge importance in propagation. Our study previously has proposed a novel centrality and analyzed its characteristics by comparing it with other centralities [4]. Our centrality index of an edge is defined by the influence on propagation when a link corresponding to the edge is removed. The activity propagation is modeled by an equalization process of exchanging variable amount among vertices.

This paper investigates how the information diffusion can be suppressed by using the proposed centrality. Specifically, the range of information diffusion on a network is compared with that on the transformed network where some edges with higher centrality are deleted. Then, we analyze how much the range of information diffusion becomes smaller after the edges are removed. The results of the numerical experiments on the proposed centrality and betweenness centrality demonstrate that it is possible to reduce the range of information spreading by deleting edges based on the proposed centrality. This fact implies that this centrality is useful as an index to prevent the information diffusion.

The rest of this paper is organized as follows, Section II defines the model that represents how people interact with each other through a network, Section III proposes the new edge centrality based on the propagation model, and Section IV investigates how the information diffusion can be suppressed by using the proposed centrality.

II. Dynamics of Propagation

A. Modeling

This section defines the model that represents how people interact with each other through a network. A given graph $G = (V, E)$ is an undirected and connected graph with vertex set V and edge set E, suppose that each vertex $v_i \in V$ has a variable $W_i(t) \in \mathbb{R}^+$ at a time $t \in \mathbb{N}$, $W_i(t)$ is referred as weight. At each time t, vertices transfer some of their weights to adjacent

 ISBN: 978-1-61208-686-6

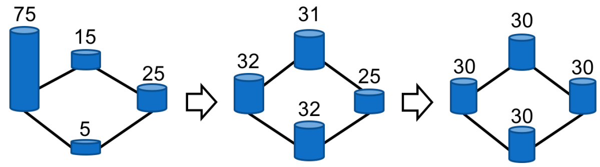


Figure 1. Dynamics of propagation.

vertices, in order to reduce the difference in weights among vertices. The weights are not lost or externally added while transferred, and the total weights of the graph as a whole are conserved (Figure 1). Hence, the sum and ensemble average of all weights is constant and independent of time, and the weight of each vertex approaches the average value over time. If the difference between the ensemble average and each weight of all vertices is less than a threshold ε, this condition says that "the weights have converged". Let the ensemble average of weights be $\langle W \rangle$, the condition of weights convergence can be written by the following equation.

$$|\langle W \rangle - W_i(t)| \leq \varepsilon \quad \forall v_i \in V. \tag{1}$$

In this model, the force trying to synchronize the states at connected vertices works. When there is no state difference between vertices, they are stable and do not mutually influence each other. The force trying to synchronize state with others is seen in various fields such as restoring force in physics, peer pressure in psychology, imitation in sociology. Thus, it is assumed that our model is a universal model describing characteristics commonly included in many models.

B. Propagation Rules

This section shows the rules of weight transfer. If the weight of a vertex is regarded as loads (works), the synchronization process of our model can be thought of as a process of solving the load balancing problem. We define the propagation rules using the simple load balancing algorithm: diffusion algorithm and local equalization algorithm.

1) Diffusion Algorithm: This algorithm determines the rule of weight transfer based on the physical diffusion process such as chemical substances [5]. Specifically, when the vertex v_i transfers some weight to adjacent vertices v_j, the amount of transferred weight can be calculated by the difference between $W_i(t)$ and $W_j(t)$ multiplied by a nonnegative constant A_{ij}. A_{ij} is called the diffusion coefficient and can be interpreted as a parameter representing the ease of transferring weights. When the weight of each vertex at a time t is $W_i(t)$, the weight $W_i(t+1)$ at the next time can be written as follows.

$$W_i(t+1) = W_i(t) + \sum_{v_j :\, neighbor\ of\ v_i} A_{ij}(W_j(t) - W_i(t)). \tag{2}$$

Boillat proposed (3) as a method of choosing A_{ij}. $\deg(v_i)$ expresses the degree of v_i. In the experiment, we assumed the homogeneity of the link for simplicity and set any element of A_{ij} to the inverse of one plus the maximum degree of vertices.

$$A_{ij} = \frac{1}{1 + \max\{\deg(v_i), \deg(v_j)\}} \qquad v_i, v_j \in V. \tag{3}$$

The propagation process based on this rule represents that each node is gradually affected by the surroundings and the influence spreads.

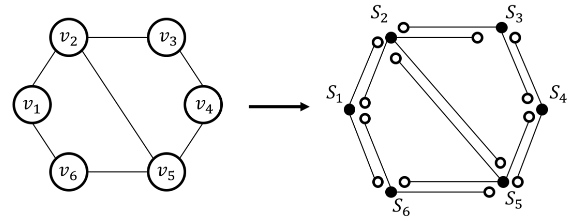


Figure 2. Extracting sub-star graphs.

2) Local Equalization Algorithm: This algorithm repeats the operation of converging the weights locally for each subgraph [4]. First, a sub-star graph S_i centered on each vertex v_i is extracted from the graph G. The sub-star graph S_i is a subgraph induced by one vertex v_i and its adjacent vertices (Figure 2). Since sub-star graphs are extracted for each vertex, the number of elements of the set of sub-star graphs S extracted from the graph G is equal to $|V|$. Next, select the sub-star graph S_i in random order from the sub-star graph set S, and locally converge the weights of the vertices included in S_i. When the local convergence of all sub-star graphs in S is completed, the time t goes to the next. The weight of the vertex v_k included in S_i changes as follows, when converging locally on the sub-star graph S_i.

$$W_j(t+1) = \frac{1}{|S_i|} \sum_{v_k \in S_i} W_k(t) \qquad v_j \in S_i. \tag{4}$$

Unlike the diffusion algorithm, this rule represents the propagation that is very susceptible to mutual influence, because the states are instantaneously synchronized at the surroundings around a node.

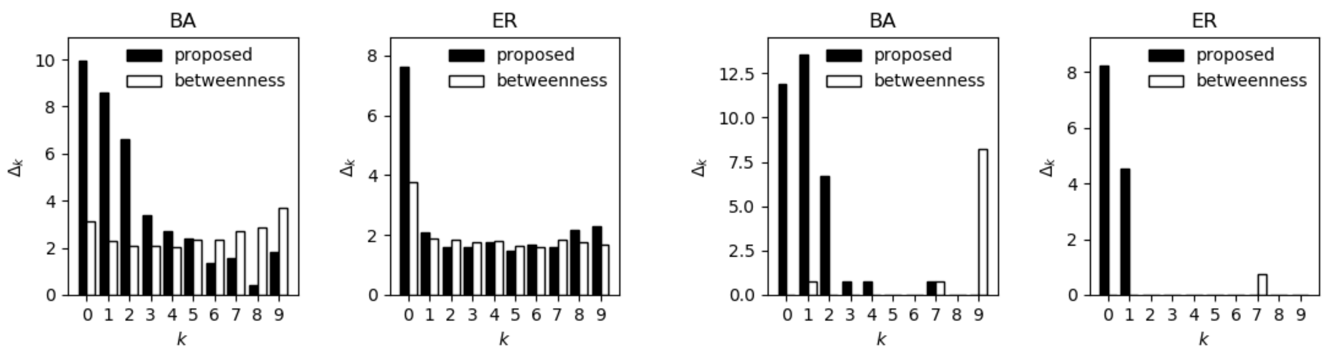
III. Edge Centrality Proposal

Using the model of Section II-A, we define the centrality of edge. Here, the time until the convergence of the weights is used as an index of the centrality. In order to define the importance of the edge e, converge the weights of the graph G' obtained by removing an edge e from the graph G. At that time, the weights transmitted along the edge e has to be propagated bypassing e, thus the weights would be difficult to converge. Therefore, the importance of an edge is defined by the difficulty of weights convergence when the edge was removed. Specifically, the importance of the edge e is defined by the ratio of the time to the convergence of the original graph G and the graph G' removed e. The time until convergence depends on the weights $\boldsymbol{W}_0 = \{W_1(0), \ldots, W_n(0)\}$ at the initial ($t = 0$). Hence, the importance is calculated with the initial weights of several patterns, and the average is used as the centrality value. Let $\mathcal{W}_0$ be a pattern set of the initial weights and $T(G, \boldsymbol{W}_0)$ be the time until convergence, the importance $D(e)$ can be written as follows.

$$D(e) = \frac{1}{|\mathcal{W}_0|} \sum_{\boldsymbol{W}_0 \in \mathcal{W}_0} \frac{T(G', \boldsymbol{W}_0)}{T(G, \boldsymbol{W}_0)}. \tag{5}$$

From this definition of importance, if the graph is not connected when an edge is removed, there is a possibility that the weight cannot converge in the propagation model. Therefore, the importance $D(e)$ cannot be defined unless a graph is at least two connected.

 ISBN: 978-1-61208-686-6



(a) IC model. (b) LT model.

Figure 3. The value of Δ_k for each k, graph models, and information diffusion models.

IV. Experiments

The proposed centrality defines the edge importance by measuring how much the propagation becomes more difficult when an edge is removed. Therefore, it is assumed that the propagation on the network can be suppressed by deleting the edges with high importance in this centrality. In addition, the propagation process expressed by the diffusion algorithm (in Section II-B1) can be used for analysis of the information diffusion [6]. This experiment analyzes how much the range of information diffusion becomes smaller after the edges are removed. These removed edges are chosen in descending order of the importance in our centrality using the diffusion algorithm as the propagation rule.

A. Simulation Setting

1) Information Diffusion Model and Parameter Setting: In order to simulate the diffusion process of information, this experiment uses two popular information diffusion models: Independent Cascade (IC) model [7] and Linear Threshold (LT) model [8]. In these models, each vertex has either *active* or *inactive* state, and active represents a state receiving the information and inactive represents a state which has not yet been received the information respectively. Also, the information diffusion on the network is expressed as the increase of active vertices. In the IC model, the set of active vertices at the beginning and diffusion probability $p_{u,v}$ for each edge (u, v) are given, and active vertex u conveys the information with probability $p_{u,v}$ to its adjacent vertex v. On the other hand, the LT model sets the set of active vertices at the beginning, the weight $\omega_{u,v}$ for each edge (u, v), and the threshold θ_v of each vertex v as initial parameters. The inactive vertex v is affected by all adjacent active vertices u according to $\omega_{u,v}$, and v becomes active when the following equation is satisfied.

$$\sum_{(u,v)\,\in\, E \text{ s.t. } u \text{ is active}} \omega_{u,v} \geq \theta_v \tag{6}$$

Assuming the homogeneity of nodes and links, this experiment set the diffusion probability $p_{u,v} = 0.2$ in the IC model, the weight of edge $\omega_{u,v}$ equals the reciprocal of $\max\{\deg(u), \deg(v)\}$ and the threshold $\theta_v = 0.25$ in the LT model [9]. In addition, the number of active vertices at the beginning is set to $0.25|V|$ for both models.

2) Definition of Indicators for the Information Diffusion Suppression: We simulate the information diffusion by the IC or LT model on a graph $G = (V, E)$, and define $A(G)$ by the expected number of active vertices at the end of that diffusion process. The removed edges are decided as the following. First, we rearrange the element of E by the descending order of the centrality value, that is, the importance of e_i is larger than that of e_j if i is smaller than j. E_k is defined as E divided into $|E|/l$ sets, as in the following equation.

$$E_k = \{e_{kl}, e_{kl+1}, \ldots, e_{kl+l-1}\}, \quad k \in \{0, \ldots, |E|/l - 1\} \tag{7}$$

According to this equation, E_0 is composed of the edges with the highest centrality value. Let G_k be the graph obtained by removing edges included in E_k from G, run the information diffusion model on each G_k to calculate $A(G_k)$. Then, we define $\Delta_k = A(G) - A(G_k)$ as an indicator to how much the information diffusion can be suppressed. This experiment uses the average value obtained by running the diffusion simulation for 1000 times as $A(\cdot)$, and set $l = 0.1|E|$ to investigate the effect of removing the 10 percent edges from the whole.

3) Graph Topology: Graphs generated based on Erdős-Rényi (ER) model [10] and Barabási-Albert (BA) model [11] were used for experiments. The ER model is the simplest random graph. This model generates edges with probability p between arbitrary two vertices. This experiment determines the number of vertices is 180 and the edge generation probability $p = 0.1$. Also, we add some edges randomly to the graph generated by the ER model in order to satisfy the two connectivity, because the proposed centrality can not be defined on the two connected graphs. The BA model generates a graph with scale-free. Many networks in the real world have been reported to be scale-free, for example, social networks and the world wide web. The BA model evolves a graph to add vertices and edges randomly by repetition. In the experiment, the number of vertices of the initial graph is 10, and every time a new vertex is added to the existing graph, 10 new edges are added to that. We add a new vertex until the number of vertices equals 180.

4) Parameter for the Proposed Centrality: Since the value of the proposed centrality varies depending on the initial weight of vertices, we use the average value calculated by 100 samples of the initial weights as the centrality value (that is $|\mathcal{W}_0| = 100$). In addition, as described above, the propagation rule is according with the diffusion algorithm, and the diffusion coefficient A_{ij} of an edge is set by (3).

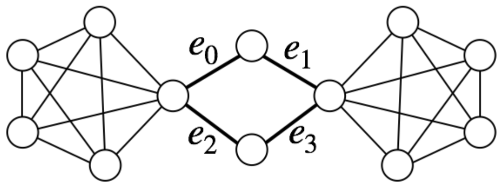


Figure 4. Example of a graph that can not efficiently suppress information diffusion with the proposed centrality

B. Results and Discussion

Figure 3 is a bar chart of the value of Δ_k for each k, graph model, and information diffusion model. For example, the left of Figure 3a represents the simulation result on information diffusion using IC model on a graph generated by the BA model. The smaller the value of k is getting, the more edges with higher centrality values are removed. The experiment uses the proposed centrality using the algorithm in Section II-B1 as a propagation rule and the betweenness centrality [3] for the comparison.

All the charts clearly show that Δ_0 of the proposed centrality is largest in all combinations of information diffusion models and graph models. The number of active vertices after the information diffusion simulation is reduced by elimination of an edge with the higher value in the proposed centrality. This fact suggests that the proposed method is useful to suppress information diffusion as an indicator in comparison with the betweenness centrality.

On the other hand, this result shows that Δ_1 also takes a large value. This fact suggests that there is a possibility that the value of Δ_0 still can be increased by changing the edges in E_k. According to (5), the proposed centrality is defined based on the influence when a single edge is removed, and measures the function of a single edge. Therefore, it is considered difficult to evaluate the influence of removing multiple edges simultaneously like the experiment. Take the case that we remove two edges from the graph in Figure 4 (that is $l = 2$). Let the four edges drawn with bold be the edges with the highest centrality value. In this case, we can divide it into the unconnected graph by removing e_0 and e_2, so it is assumed that the information diffusion can be effectively suppressed. When the two removed edges are chosen based on the centrality value to suppress the information diffusion, since the centrality values of the four bold edges are the same, not $\{e_0, e_2\}$ but $\{e_0, e_1\}$ might be chosen, it is not optimal.

We try to extend the definition of edge importance to consider multiple edges. In fact, it is unlikely that the situation of removing only one edge (corresponding to the link of such as human relations) occurs. Therefore, it is important to consider the influence of removing multiple edges, when applying this method to information diffusion and other propagation phenomena in the real world. In Section III, G' is defined by G without an edge e, and the importance of an edge e is calculated depending on the propagation process on G'. We change the G' in (5) like the following equation in order to calculate the influence of removing multiple edges in a subset E' of E.

$$D(e) = \frac{1}{|\mathcal{W}_0|} \sum_{\boldsymbol{W}_0 \in \mathcal{W}_0} \frac{T(G \setminus E', \boldsymbol{W}_0)}{T(G, \boldsymbol{W}_0)}. \tag{8}$$

This equation measures the influence that the edges in E' are removed from a graph G, so it can evaluate the importance of the multiple edges. However, the number of combinations to select a subset of edges is enormous, and it is necessary to devise an efficient method for calculating them.

V. Conclusion and Future Works

This paper investigates how the information diffusion can be suppressed by using the proposed centrality. The results of the numerical experiments on the proposed centrality and betweenness centrality demonstrate that it is possible to reduce the range of information spreading by deleting edges based on the proposed centrality. We also extend the definition of edge importance to consider the influence for the propagation of multiple edges.

There are future tasks. First, we use only the two type of the ER model and BA model for the experiment, so it is necessary to simulate the real social network topology in the future. In addition, we will analyze the effect when changing parameters for such as graph models, information diffusion models, and the ratio of removing edges, and also add other centralities for comparison. Then, this paper investigated whether the proposed method can be applied to only the information diffusion. However, it is assumed that the proposed centrality can be widely used for analysis of not only for the information diffusion but also the dynamic propagation processes. Therefore, for example, we will investigate whether this method can be applied to such as efficient control of power-flow network for smart grid and design of the communication protocol on mobile ad hoc network.

References

[1] A. Guille, H. Hacid, C. Favre, and D. A. Zighed, “Information diffusion in online social networks: A survey,” SIGMOD Rec., vol. 42, no. 2, Jul. 2013, pp. 17–28.

[2] T. E. Ulrik Brandes, Network Analysis: Methodological Foundations. Springer-Verlag Berlin Heidelberg, 2005.

[3] L. C. Freeman, “A set of measures of centrality based on betweenness,” Sociometry, vol. 40, no. 1, 1977, pp. 35–41.

[4] K. Hamda, G. Ishigaki, Y. Sakai, and N. Shinomiya, “Significance analysis for edges in a graph by means of leveling variables on nodes,” in Proceedings of the 7th International Conference on Computing Communication and Networking Technologies, ser. ICCCNT ’16. New York, NY, USA: ACM, 2016, pp. 27:1–27:5.

[5] J. E. Boillat, “Load balancing and poisson equation in a graph,” Concurrency and Computation: Practice and Experience, vol. 2, no. 4, 1990, pp. 289–313.

[6] E. Takazawa, K. Hamada, and N. Shinomiya, “Characteristic analysis of the edge centrality based on activity propagation in information networks,” in Proceedings of 2018 IEEE International Conference on Systems, Man, and Cybernetics, 2018, pp. 1578–1582.

[7] J. Goldenberg, B. Libai, and E. Muller, “Talk of the network: A complex systems look at the underlying process of word-of-mouth,” Marketing letters, vol. 12, no. 3, 2001, pp. 211–223.

[8] M. Granovetter, “Threshold models of collective behavior,” American Journal of Sociology, vol. 83, no. 6, 1978, pp. 1420–1443. [Online]. Available: https://doi.org/10.1086/226707

[9] D. Kempe, J. Kleinberg, and E. Tardos, “Maximizing the spread of influence through a social network,” in Proceedings of the Ninth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, ser. KDD ’03. New York, NY, USA: ACM, 2003, pp. 137–146. [Online]. Available: http://doi.acm.org/10.1145/956750.956769

[10] P. Erdos and A. Rényi, “On the evolution of random graphs,” Publications of the Mathematical Institute of the Hungarian Academy of Sciences, vol. 5, no. 1, 1960, pp. 17–60.

[11] A.-L. Barabási and R. Albert, “Emergence of scaling in random networks,” Science, vol. 286, no. 5439, 1999, pp. 509–512.