How to maximize the influence to the whole network

Zesen Zhang

Abstract— Social behaviors and choices spread through interactions and may lead to a cascading behavior. Understanding the way social cascades spread in a network is crucial for many applications ranging from viral marketing to political campaigns. The behavior of cascade depends crucially on the model of cascade or social influence and the topological structure of the social network.

In this paper we try to learn the necessary number of initial seeds that need to be influenced in the Preferential-Attachment model and Erdós-Rényi model. Further, we try to attach the evolving character into the network which can make our results more reliable. Though, there are many different kinds of distribution model we choose the k-complex contagion model to be our distribution model, which has been proved to be more close to the reality. What's more, we try to step our work into a more general model that we cascade the parameter k into the specific distribution.

I. INTRODUCTION

As we are in the information era and the social network has became much larger recently. With the development of the internet, information spread much quicker than ever before. And human activity is embedded in a network of social interactions, which can spread information, beliefs, diseases, technologies, and behaviors. Therefore, a better understanding of the social interaction and social network behavior can help us learn how the information diffused in the network. Also, learning all these character in the network can help the business men or governors find the best way to advertise their products[2], [3] or diffuse their politics[1].

There are two important factors in determining the scope and rate of such diffusion: the model of contagions, i.e., how a node is influenced by its neighbors; and the network topology. We discuss these two factors separately.

As for the Social Contagion Models. The study of contagions starts from the study of infectious diseases and epidemics[4]. Soical behaviors and decisions are "contagious" too. And there are many different kinds of contagion model in the recent work. For example, there are independent cascade model, linear threshold model, k-complex contagion model and so on.

We call contagions simple when the influence g_v is submodular-that is $g_v(S^{'} \cup \{x\}) - g_v(S^{'}) \leq g_v(S \cup \{x\}) - g_v(S)$, if $S \subset S^{'}$–and call contagions complex when this fails to hold. In a complex contagion, there could be an initial barrier such that no activation is possible until the barrier is crossed. If we define $f(S)$ as the expected number of infected nodes when the vertices in S are chosen as the initial seeds, then if g_v is submodular for all nodes, then f is submodular as well[5].

The complex contagion model contains the monotonicity and submodularity character. However, work done on complex contagions is much more limited and so far focused on a simplistic single threshold model called k-complex contagion model. In k-complex contagion model, all nodes have the same threshold k. A node becomes infected if and only if at least k of its neighbors have been infected. It has been shown that a k-complex contagion is generally slower and more delicate than simple contagion k=1[7], [8], [9].

One of the limitations of this k-complex contagion model is the dependency on the fixed threshold k for all nodes in the network. In practice there are people who like to try out new things and are more risk driven while others are risk averse. Therefore the threshold function is not necessarily uniform. Therefore, we try to step our work further to assume the parameter k distributed in a specific distribution which will make the result more reliable.

In addition to the model of cascade, the model of network is also important. A lot of mathematical models have been developed to capture some of the attributes of real world social networks. A celebrated set of results are the family of small world graphs[10], [11], [12] and the family of graphs that produce power law degree distribution[13], [14], [15]

In this work, we examine an evolving social network based on the Erdős–Rényi model and Preferential Attachment model. The Erdős–Rényi model is a model with some simple but useful characters. It is a classics model in the social network theory. In the evolving Erdős–Rényi model, nodes arrive in a sequential order. And each edge attached to the former edge with probability p. This is an interesting model that the later coming edge supposed to emit more nodes. And in the Preferential Attachment models, nodes arrive in a sequential order. Each node chooses m edges from the nodes that arrive earlier. When an edge is added, the neighbor is selected with probability proportional to its current degree. This model generates graphs with a power law degree distribution and has been used to explain the observations in web graphs and social networks. Therefore, we can see that both model has some fantastic character and we can further detect our results based on such characters.

As for our results, we first detect the evolving character based on the Erdős–Rényi model. In this model we find out that the character of the evolving network can help us extend the influence of the original network, even the later one does not bring any influence into the network.

This result matches our general sense that a network in a critical state. A suddenly coming in man may break this state and spread the influence to the whole network.

And for the Preferential Attachment model, we find out that we only need to influence the h number of in the first e^h scale of network can we diffuse the influence to the whole network. This is a more general results than the former work who are going to influence the first k number of nodes to influence the whole network.

Further, we extended our results into a more general contagion model. We try to generalize the parameter k into some distributions which will make our results closer to the reality.

II. PRELIMINARIES

A. k-complex contagion model

We define a k-complex contagion process in an undirected graph, where k=O(1). We assume that we are given a graph G, which might have been generated by an evolving process.

Definition 1. Given a graph G, a k-complex contagion $CC(G, k, p)$ is a contagion that initially infects vertices with probability p, initial seeds, and spreads over graph G. The contagion proceeds in rounds. At each round, each vertex with at least k infected neighbors becomes infected. We are interested in the necessary number of initial nodes for all the nodes to be infected.

B. Erdős–Rényi Model

We first proceed our analysis on the Erdős–Rényi model. It is the most useful random model and has been used in many different kinds of works. We try to involve the evolving character into this model which will make our results closer to the realistic.

Definition 2. In the $G(n, p_0)$ model, a graph is constructed by connecting nodes randomly. Each edge is included in the graph with probability p_0 independent from every other edge. And in the evolving model, each node comes in sequentially. Each edge attached to the former nodes with probability p_0. Therefore, the n_{th} node emits np_0 number of edges as expected.

C. Preferential Attachment Model

Preferential Attachment model is a model that nodes are supposed to connect to the nodes with higher degree. Therefore, the PA model makes the network to be polynomial which has been proved to be the typical character of the social network.

Definition 3. We define the independent preferential attachment model, $PA_m(n)$: We start with a complete graph on m+1 nodes. At each subsequent time step t=m+2,...,n a node v arrives and adds m edges to the existing vertices in the network. Denote the graph containing the first n-1 nodes as G_{n-1}. For each new vertex, we choose $w_1, w_2, ..., w_m$ vertices, possibly with repetitions from the existing vertices in the graph. For each i, w_i is selected from the set of vertices of G_{n-1}

with probability proportional to the vertices' degree in G_{n-1}. Then we draw edges between the new vertex and the w_i's. Repeated w_i's cause multiple edges. Note that $\sum_{v \in V(G_n)} deg(v) = 2mn$

III. Erdős–Rényi Model

In this section, we analyze the number of nodes that need to be infected to spread the influence to the whole network based on the Erdős–Rényi Model. For the discussion below, we first analyze a specific character of the evolving Erdős–Rényi model and then we use the Markov process and the Chernoff bound to make our proof.

As we can see from the following figure 1, the former network contains some infected nodes and some uninfected nodes. We assume that the contagion model is the 2-complex contagion model. As we can see there are no nodes that can be infected in the left figure, therefore, the diffusion of the whole network stopped. However, in the next time period, an uninfected nodes come in and it connects to every node with probability p_0. Therefore, we can assume that the node luckily connect to the two infected nodes and connect to another uninfected node. Surprisingly, as we can see from the figure on the left, this node give a "drive" to the whole network and revitalize the whole network. Because of this node, the influence can be diffused into the whole network. And we can call this phenomenon as "avalanche" phenomenon.
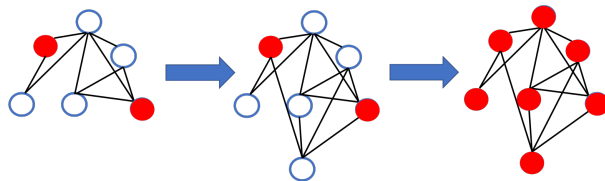

Fig. 1. This is an illumination that an uninfected node can revitalize the influence function in the whole network

Based on such character, we want to explore the how the evolving character can make difference in the diffusion of the influence.

A. Main results for Erdős–Rényi Model

We would like to first state the main theorem that characterizes the behavior of k-complex contagion model on the Erdős–Rényi graph $G(n, p_0)$

Theorem 3.1. Let $G(n, p_0, p, t_0)$ be the Erdős–Rényi graph that node connects to the former node with probability p_0. And the proportion that "infected" nodes have in the first n nodes is p. And latter it will come in with t nodes who are not initial nodes. And our results can be seen as follows:

if $np_0p + \frac{t_0 p_0 p(t_0 + 2n + 3)(t_0 - 2)}{(1 - p_0 p)^2 * 2} \geq k * lnn$ then the influence can be diffused to the whole network with probability $1 - \Theta(\frac{1}{n})$. While $n \to \infty$, the $\frac{1}{n} \to 0$. We can say the whole network will be influenced.

B. Proof of Theorem 3.1.

First we want to use the Markov process to portray the "avalanche" phenomenon. We simply define the I_t to be the state that whether node t has been infected. If $I_t = 1$, we can say the node has been influenced. Otherwise if $I_t = 0$, the node has not been influenced. And we use N_t^u to portray that node u connects to the node t. The k is the parameter of the k-complex contagion model. And we assume that the node u connects to the node t with probability q among the whole uninfected nodes. Then our "avalanche" phenomenon can be portrayed in the following equation:

$$Prob[I_t|N_t^u, I_u] = \frac{\sum_{l=k}^{n} C_n^u (p_0 p)^u (1 - p_0 p)^{n-u} (1 - (1-q)^{n-u})}{1 - (1 - (1-p_0 p)q)^n} \quad (1)$$

Then we want to find the probability that $I_u = 1$. And it is because the u is a general node that it can present the whole nodes that in the first n nodes and did not get infected. Therefore we want to prove that based on the $np_0 p + \frac{t_0 p_0 p (t_0 + 2n + 3)(t_0 - 2)}{(1-p_0 p)^2 * 2} \geq k * lnn$ all the nodes will get infected, i.e. $Prob[I_u = 1]$. Therefore, we can formulate the $Prob[I_u = 1]$ as follows:

$$Prob[I_u = 1] = Prob[\sum_{l=1}^{n+t} N_u^t * I_t \geq k]$$
$$= Prob[p_0 pn + \sum_{h=n+1}^{n+t_0} Prob[I_t|N_u^t, I_u] \geq k] \quad (2)$$

And with the definition of the $G(n, p_0, p, t_0)$ we can find out that the $(1 - p_0 p) * q = p_0$. Therefore we can simplify our equation in the following way:

$$Prob[I_u = 1]$$
$$= Prob[p_0 pn$$
$$+ \sum_{h=n+1}^{n+t_0} \frac{\sum_{l=k}^{t} C_t^u (p_0 p)^u (1 - p_0 p)^{t-u} (1 - (1-q)^{t-u})}{1 - (1 - p_0)^t}] \quad (3)$$

As for the $Prob[I_t|N_t^u, I_u]$ we can use the Taylor formula $(1-p)^q = 1 - pq \ while \ p \to 0$ to simplify our results. And we can simplify the results in the following way:

$$Prob[I_t|N_t^u, I_u] = \frac{\sum_{l=k}^{t} C_t^u (p_0 p)^u (1 - p_0 p)^{t-u-1}(t-u)}{t} \quad (4)$$

Then we choose to unfold the binomial coefficient and reform them into two parts and as $p_0 p$ is a very small number we ignore the two order formula of $p_0 p$ and simplify the results as follows:

$$Prob[I_t|N_t^u, I_u] =$$
$$\frac{p_0 p}{t} [\frac{(t+1)^2}{(1-p_0 p)^2} - \frac{t-1}{t}] \quad (5)$$

Then we take our results into the formula of $Prob[I_u = 1]$ then we have following formula:

$$Prob[I_u = 1]$$
$$= Prob[p_0 pn + \sum_{h=n+1}^{n+t_0} (\frac{p_0 p}{t} [\frac{(t+1)^2}{(1-p_0 p)^2} - \frac{t-1}{t}]) \geq k]$$
$$= Prob[p_0 pn + \frac{p_0 p}{(1-p_0 p)^2} * \frac{(t_0 + 2n + 3)(t_0 - 2)}{2} \geq k]$$
$$= 1 - Prob[p_0 pn + \frac{p_0 p}{(1-p_0 p)^2} * \frac{(t_0 + 2n + 3)(t_0 - 2)}{2} < k]$$
$$(6)$$

Then we use the chernoff bound to reform our results and we can see that the $Prob[I_u = 1]$ can be reformed into the following way:

$$Prob[I_u = 1]$$
$$= 1 - \frac{exp(t_1 k)}{exp(t_1 * (p_0 pn + \frac{p_0 p}{(1-p_0 p)^2} * \frac{(t_0 + 2n + 3)(t_0 - 2)}{2}))} \quad (7)$$

Then we can see that once $np_0 p + \frac{t_0 p_0 p (t_0 + 2n + 3)(t_0 - 2)}{(1-p_0 p)^2 * 2} \geq k * lnn$ the $Prob[I_u = 1]$ will be equal to $1 - \Theta(\frac{1}{n})$. Once $n \to \infty$ the $Prob[I_u = 1]$ tends to 1.

IV. Preferential Attachment Model

The Preferential Attachment Model has many interesting results. Take the degree of the node for example. The [15] shows that the preferential attachment model is a scale-free power-law distribution and this feature is found to be a consequence of the two generic mechanisms that networks expand continuously by the addition of new vertices, and new vertices attach preferentially to already well connected sites. Therefore, more and more researchers try to detect the results based on the Preferential Attachment Model which seems to be closer to the reality.

Therefore, we would like to detect the evolving character in the PA model. We try to use the Markov process to portray every step of the diffusion of the influence and finally find out the number of initial nodes that we need to spread the influence to the whole network.

A. Markov chain

First we would like to give an eye to the Markov process which will make our later illustration easier.

Assume that node u is the $i - th$ node in the arrival order in G. Let V_{i-1} be the set of first $i - 1$ nodes in G and X_{i-1} be the set of infected nodes in V_{i-1}. If u's threshold is $R_u = k$, u is infected if and only if among the m edges u issues, at least k of them land in nodes in X_{i-1}. Now consider a specific edge of u,

we define Y_i as the probability that this edge lands in an infected node(e.g. in X_{i-1}). Y_i depends on the attachment rule A and the set of nodes that are infected so far. For example, if the edges of u are preferentially attached, i.e. with probability proportional to the current degree of the nodes, Y_i is the ratio of the infected degree $Y_i = \sum_{v \in X_{i-1}} deg(v) / \sum_{w \in V_{i-1}} deg(w)$. Here $deg(v)$ is the total degree of each node v (counting both incoming and outgoing edges).

Next we can compute the probability of node u being infected when its threshold is $R_u = k$. For that to happen, among the m edges of u, at least k of them need to land on a node in X_{i-1}.

$$Prob\{Infection\ of\ u | R_u = k\} = \sum_{l=k}^{m} C_l^m Y_i^l (1-Y_i)^{(m-l)} \tag{8}$$

Therefore, the random process $\{Y_t : t = m+1, ..., n\}$ in $SA_M(n)$, is a Markov chain that only depends on the previous state of the process.

B. Main result of the Preferential Attachment Model

Similar to the former pattern, we first show our main result.

Theorem 4.1. We only need to randomly choose h number of nodes in the first e^h number of nodes that we are going to spread our influence to the whole network. The results can also be seen in the following figure2
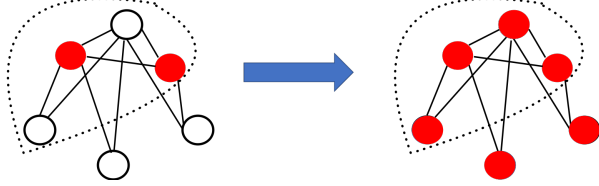

Fig. 2. This is an illumination that we only need to randomly choose h number of nodes in the first e^h number of nodes.

C. Proof of Theorem 4.1.

The prove of the Theorem 4.1. can be divided into two parts. First we would like to calculate the "avalanche" point of the PA model. We assume that when the proportion of the infected nodes reaches to p is the "avalanche" point of the PA model. Then our results can be seen in the following equation:

$$Prob[I_u = 1] = Prob[\sum_{l=1}^{n+t} N_u^t * I_t \geq k]$$
$$= Prob[2mnp \geq k] \tag{9}$$
$$= 1 - Prob[2mnp < k]$$

Then we can use the chernoff inequality to reform the equation into the following way:

$$1 - Prob[2mnp < k] = 1 - \frac{e^{t_1 k}}{e^{2mnp}} \tag{10}$$

Therefore we find out that we only need to satisfy the $p = \frac{lnn}{n}$ that we can make the $Prob[I_u] = 1 - \Theta(\frac{1}{n})$. As $n \to \infty$ the $Prob[I_u = 1] \to 1$.

However, the proportion p was just what we assumed. In the PA model, the influence may spread before it stops. Therefore, maybe we are going to need less initial seeds to spread our influence. Therefore, we assumed that we need to put our initial seed into the network with probability θ to make the network stops with proportion p nodes get infected. We try to use the Markov chain to do our analysis. Therefore we have the following equation:

$$Y_t - Y_{t-1} = \frac{2mY_{t-1} + mp_{sum} + \gamma}{2tm} \tag{11}$$

The p_sum is the probability that the edge of new node to be the "infected" edge. And it can be shown in the following way:

$$p_{sum} = \theta + (\sum_{k=k_0}^{m} C_m^k Y_{t-1}^k (1 - Y_{t-1})^{m-k})(1 - \theta) \tag{12}$$

We assume the h to be the $max_{k_0 \leq k \leq m} C_m^k$. Then we can simplify the p_{sum} into the following way:

$$p_{sum} \leq \theta + (h * Y_{t-1}^{k_0} + h \frac{Y_{t-1}^{k_0}}{(1 - Y_{t-1})^2} - mh Y_{t-1}^{k_0+1})(1 - \theta) \tag{13}$$

And the γ is the probability that the new node make the former node become infected and make their edge become "infected". Therefore it can be seen in the following equation:

$$\gamma = \frac{1}{2m(t-1)} \frac{1}{(k_0 - 1)!} Y_{t-1}^{k_0-1} \sum_{u=1}^{t-1} \frac{d(u) * d(u)!}{(d(u) - k + 1)!} \tag{14}$$
$$* [1 - [d(u) - k + 1] Y_{t-1}](1 - \theta) p_{sum}$$

$d(u)$ is the degree of the node u and it can be calculated in the following way:

$$\frac{\partial d(u_i)}{\partial t} = \frac{d(u_i)}{2t}$$
$$d(u_i) = (\frac{t}{t_i})^{\frac{1}{2}} * m \tag{15}$$

We let the $I = \frac{k_0 - 2}{(k_0 - 1)!}$, $E = exp\{((k_0 - 1) - (\frac{1}{2})^{\frac{k_0+1}{2}} * \frac{k_0-2}{k_0-1} * t^{\frac{1}{2}}) Y_{t-1}\}$. Then we can simplify the γ into the following equation:

$$\gamma \leq I * E * \frac{t^{\frac{k_0}{2}}}{t-1} Y_{t-1}^{k_0-1} (1 - \theta) p_{sum} \tag{16}$$

Therefore the $Y_t - Y_{t-1}$ can be simplify into the following way:

$$Y_t - Y_{t-1} \le [\frac{1}{2t} + \Theta((\frac{lnt}{\sqrt{t}})^{k_0})]\theta + [\Theta((\frac{lnt}{t})^{k_0}) + \Theta((\frac{(lnt)^2}{t^{\frac{3}{2}}})^{k_0}]$$
$$(17)$$

When $k_0 \ge 3$, $(\frac{lnt}{\sqrt{k}})^{k_0} = o(\frac{lnt}{t})$. Then we can use the union bound and let $c = \Theta(\theta^2)$. Therefore, we have the following equation:

$$Prob((Y_t - Y_0) \le t) \le exp\{-\frac{t^2}{2c}\}$$
$$Prob(Y_t - \theta \le \frac{lnt}{t} \le exp\{\frac{-(\frac{lnt}{t})^2}{2\Theta(\theta^2)}\}$$
$$(18)$$

As $\theta = o(\frac{lnt}{t})$, $exp\{\frac{-(\frac{lnt}{t})^2}{2\Theta(\theta^2)}\}$, Y_t is an increasing function. Therefore we have $\theta = \Theta(\frac{lnt}{t})$.

While the $k = 2$, we have another equation to formulate the pattern. The equation can be seen in the follow:

$$E[Y_t] = \frac{2(t-1)mY_{t-1} + mp_{sum} + \gamma}{2tm} \quad (19)$$

We insert the p_{sum} and the γ into the equation and we can reform it into the following way:

$$E[Y_t] = \frac{Y_{t-1}}{2m(t-1)}\theta(1-\theta)[m^2(t-1)lnt - [m^3(t-1)^{\frac{3}{2}}$$
$$* \frac{1}{2}[1 - (\frac{1}{t-1})^{\frac{1}{2}}] - m^2(t-1)lnt]Y_{t-1}]$$
$$(20)$$

In order to make our results easier to read we use A, B, C, P, Q on behalf of some equations:

$$A = -(\frac{1}{8}\frac{m(t-(\frac{1}{2})^{\frac{1}{2}}}{t}\theta(1-\theta) - \frac{1}{8}\frac{m}{t}\theta(1-\theta) - \frac{1}{4}\frac{lnt}{t}\theta(1-\theta)$$
$$- \frac{m(m-1)}{2t})$$
$$B = \frac{t-1}{t} + \frac{1}{4}\frac{lnt}{t}\theta(1-\theta)$$
$$C = \frac{-\theta}{2t}$$
$$P = A$$
$$Q = -\frac{4AC - B^2 + 2B}{4A}$$
$$a_1 + \frac{1}{a_1} = \frac{-8AY_1 + 4B}{4AC - B^2 + 2B}$$

Then we have Y_t to be expressed in the following way:

$$Y_t = (\frac{8A^2}{4AC - B^2 + 2B})^{t-1}Q - \frac{1}{2}a_1^{2t-2}$$
$$+ (\frac{4AC - B^2 + 2B}{8A^2})^{t-1}Q \quad (21)$$
$$- \frac{1}{2}(\frac{1}{a_1})^{2t-2} - \frac{B}{2A}$$

Then we have the equation:

$$\frac{1}{2A} - (\frac{1}{8A^2})^{t-1}\frac{1}{4A}(\frac{1}{a_1})^{2t-2} = \frac{lnt}{t} \quad (22)$$

Therefore we can have that $\theta = \Theta(\frac{lnt}{t})$. Therefore, the theorem 4.1. has been proved.

V. CONCLUSIONS

Based on all the results we have made above. We found out that the evolving character can bring lots of interesting benefits to the results. And we believe these results will give a further step in this area.

APPENDIX

Appendixes should appear before the acknowledgment.

ACKNOWLEDGMENT

The preferred spelling of the word ÒacknowledgmentÓ in America is without an ÒeÓ after the ÒgÓ. Avoid the stilted expression, ÒOne of us (R. B. G.) thanks . . .Ó Instead, try ÒR. B. G. thanksÓ. Put sponsor acknowledgments in the unnumbered footnote on the first page.

References are important to the reader; therefore, each citation must be complete and correct. If at all possible, references should be commonly available publications.

References

[1] Lada A Adamic and Natalie Glance. 2005. The political blogosphere and the 2004 US election: divide they blog. In Proceedings of the 3rd International Workshop on Link discovery. 36-43

[2] Abhijit Banerjee, Arun G Chandrasekhar, Esther Duflo, and Matthew O Jackson. 2013. The diffusion of microfinance. Scinece 341,6144(2013)

[3] J.Coleman, E. Katz, and H.Menzel. 1957. The diffusion of an innovation among physicians. Sociometry 20(1957), 253-270

[4] Matthew O.Jackson. 2008. Social and Economic Networks/ Princeton University Press, Princeton, NJ,USA.

[5] Elchanan Mossel and Sebastien Roch. 2010. Submodularity of Influence in Social Networks: From Local to Global. SIAM J/Comput/ 39.6(2010).2176-2188.

[6] J. Wang, ÒFundamentals of erbium-doped fiber amplifiers arrays (Periodical styleÑSubmitted for publication),Ó IEEE J. Quantum Electron., submitted for publication.

[7] Roozbeh Ebrahimi, Jie Gao,Golnaz Ghasemiesfeh, and Grant Schoenebeck.2014.How Complex Contagions Spread Quickly in the Preferential Attachment Model and Other Time-Evolving Networks. arXiv preprint arXiv:1404.2668 (2014).

[8] Roozbeh Ebrahimi, Jie Gao, Golnaz Ghasemiesfeh, and Grant Schoenebeck.2015. Complex Contagions in Kleinberg's Small World Model. In Proceedings of the 2015 Conference on Innovations in eoretical Computer Science (ITCS). 63–72.

[9] Golnaz Ghasemiesfeh, Roozbeh Ebrahimi, and Jie Gao. 2013. Complex contagion and the weakness of long ties in social networks: revisited. In Proceedings of the fourteenth ACM conference on Electronic Commerce. 507–524.

[10] D.Watts, P.Dodds, and M.Newman.2002. Identity and Searchin Social Networks. Science 296(2002),1302–1305.

[11] Jon Kleinberg.2000. The small-world phenomenon: analgo-rithm perspective. In Proceedings of the 32nd annual ACM symposium on theory of Computing. 163–170.

[12] Jon M.Kleinberg, Ravi Kumar, Prabhakar Raghavan, Sridhar Rajagopalan, and Andrew S. Tomkins. 1999. The web as a graph: measurements, models, and methods. In Proceedings of the 5th annual international conference on Computing and combinatorics. 1–17.

[13] R.Kumar, P.Raghavan, S.Rajagopalan, D.Sivakumar, A.Tomkins, and E.Upfal.2000. Stochastic models for the Web graph. In Proceedings of the 41st Annual Symposium on Foundations of Computer Science (FOCS '00). 57–.

[14] Ravi Kumar, Prabhakar Raghavan, Sridhar Rajagopalan, and Andrew Tomkins. 1999.Extracting Large-Scale Knowledge Bases from the Web. In Proceedings of the 25th International Conference on Very Large Data Bases. 639–650.

[15] Si A L B, Albert R. Emergence of Scaling in Random Networks[M]// The Structure and Dynamics of Networks. 2011.