A Novel Measurement of Structure Properties in Complex Networks

Yanni Han[1], Jun Hu[1], Deyi Li[1,2], and Shuqing Zhang[1]

[1] Beihang University, Beijing 100191, China
hyn@nlsde.buaa.edu.cn
[2] Institute of Electronic System Engineering, Beijing 100039, China

Abstract. Traditional measurements provide an effective tool to study the complex large systems in the real world. These global quantities only analyze the general statistical properties and interconnectivity structure of the entire network. However the complicated interactions among the locals are indeed the origin to emergent complex behavior. So in this paper we present a new measurement to reveal the local structure properties - topology potential, which reflects the differential position of each node in the topology. It is flexible by adjusting the influence factor. We demonstrate our measurement in US politics books network. Experiments confirm that topology potential has inherently implied the traditional measurements to some extent.

Keywords: topology potential, potential distribution, complex network, degree, betweenness.

1 Introduction

The outburst of activity in the field of complex networks in recent years has been rather spectacular and amazing. And the statistical properties analysis provide an effective tool to study the complex large systems in the real world.Typically the structure of complex networks is characterized in terms of global properties [1][2], such as average shortest path length, clustering coefficient [3], assortativity [4] and other measures, especially, the degree distribution [5]. However, these global properties may sometimes fail to provide insight into the mechanisms responsible for the formation or growth of these networks. And the relationship among them is not clear. Often complex structure of networks is influenced by system-dependent local constraints on node interconnectivity [6]. Both nodes characteristics and links properties may vary over time. Thus alternative approaches that take into consideration of the local structure in real-world complex networks are therefore necessary.

In this paper we present a novel measurement-Topology Potential. It present a unified description framework for traditional measurements. The paper is organized as follows. In section 2 we give the derivation and function definition of topology potential. Section 3 we focus on the discussion of influence factor and optimization algorithm. In section 4 we do experiments on the US politics books

J. Zhou (Ed.): Complex 2009, Part II, LNICST 5, pp. 1292–1297, 2009.

network. Compared with different influence factors, we prove the reasonable of optimal factor. Furthermore, Evidences show that the new measurement has inherently implied the traditional measurements. Then conclude with Section 5.

2 Topology Potential

According to the field theory in physics, potential in a conservative field is a function of position, which is inversely proportional to distance and directly proportional to magnitude of particle's mass or charge. Inspired from the above physical idea, we introduce field into network topological structure to describe the relationship among nodes being linked by edges and to reveal the general characteristic of underlying importance distribution.

Given the network G= (V, E), V is the set of nodes, E is the set of edges and |E|=m. So topology potential can be defined as the differential position of each node in the network, that is to say, the potential of node in its position. Each node's influence will quickly decay as distance increases. Hence, we define the topological potential in the form of Gaussian function. The potential of node $v_i \in$ V in the network can be formalized as:

$$\varphi(v_i) = \sum_{j \in N} (m_j * \exp -(\frac{d_{ij}}{\sigma})^2). \tag{1}$$

Where d_{ij} is the distance between node V_i and V_j; parameter σ is used to control the influence region of each node, called influence factor; $m_i \geq 0$ is the mass of node V_i (i=1,...,n), which meets a normalization condition $\sum_{i \in N} m_i$=1.

As the modularity structure of real-world network implies that the interaction among nodes has local characteristic. In essence the topological potential score of each node can reflect nodes importance in the topology by optimizing influence factor, which can reveal the ability of each node influenced by the other nodes in the network, and vice versa.

3 Influence Factor in Topology Potential

Complex systems in nature usually behave different forms as different types of node in the topology, such as peripheral nodes, satellite connectors, global hubs, etc. Therefore position difference of nodes in topology can be regarded as the ability of interaction or influence region. We adopt two ways to obtain the influence factor of node potential. On the one hand, number of influence region can be specified by users in the context of priori knowledge. The value is ranging from 1 hop (the neighbors of each node) to the diameter of network. This way is relatively subjective and need to know some priors in advance. On the other hand, potential entropy has been introduced to measure the uncertainty of topological space, similar to the essence of information entropy. Intuitively, if each node's potential value is different, then the uncertainty is the lowest accounting for the smallest entropy. So a minimum-entropy method can be used for the optimal choice of influence factor σ. This way is more reasonable and without any pre-defined knowledge.

3.1 Definition of Optimal Influence Factor

Given a topological potential field produced by a network G=(V, E), let the potential score of each node v_1,...,v_n be $\varphi(v_1)$,...,$\varphi(v_n)$ respectively, potential entropy H can be introduced to measure the uncertainty of topological potential filed, namely

$$H = -\sum_{i \in N} \frac{\varphi(v_i)}{Z} \log\left(\frac{\varphi(v_i)}{Z}\right). \quad (2)$$

Where Z is a normalization factor. Clearly, For any $\sigma \in (0, +\infty)$, H satisfies 0≤H≤log(n) and reaches maximum value log(n) if and only if $\varphi(v_1) = \varphi(v_2)$=,...,=$\varphi(v_n)$.

3.2 Optimization Algorithm of Influence Factor σ

```
Input: Initial search range [a,b],precision threshold e;
Output:optimized s
Begin
Given sl = a+(1-t)(b-a),sr = a+t(b-a); t =
(sqrt(5)-1)/2; Hl =H(sl) and Hr = H(sr);
While |b-a|>e do
    If Hl < Hr then {
    Let b = sr, sr = sl, Hr = Hl;
    sl = a+(1-t)(b-a) and Hl = H(sl);
    }
    Else {
    Let a = sl, sl = sr, Hl = Hr;
    sr = a+ t(b-a) and Hr = H(sr);
} End while
If Hl < Hr then {s = sl}
Else {s = sr};
Return s;
End.
```

3.3 Discussion about the Range of Influence Factor

According to the properties of Gaussian function, for a given σ, its influence region approximates to $\frac{3\sigma}{\sqrt{2}}$. More details are presented in Table 1. Where $n_l(v)$ is the numbers of neighbors of node v within l hop $(1 < l < D)$ and D is the diameter of network.

Essentially topology potential and its' distribution have inherently implied traditional measurements, such as degree, closeness, clustering coefficient etc. It present a unified description framework for measurements characterization of complex networks, furthermore optimal choice of influence factor σ make the result of network analysis more reasonable.

Table 1. Range of influence factor and the formula of topology potential

Range of σ	Interaction among nodes	Formula of topology potential
$0<\sigma<\frac{\sqrt{2}}{3}$	no interaction among nodes and value of potential is 1;	$\varphi(v_i) = 1$
$\frac{\sqrt{2}}{3}\leq\sigma<2\frac{\sqrt{2}}{3}$	influence its neighbors, equivalent to the degree of nodes;	$\varphi(v_i) = 1+deg(v_i)*\exp-(\frac{1}{\sigma})^2$
$2\frac{\sqrt{2}}{3}\leq\sigma<\sqrt{2}$	influence its neighbors nodes within two hops;	$\varphi(v_i) = deg(v_i)\exp-(\frac{1}{\sigma})^2 + 1 + n_2(v_i) * \exp-(\frac{2}{\sigma})^2$
$n\frac{\sqrt{2}}{3}\leq\sigma<\frac{\sqrt{2}\min(n+1,D)}{3}$	influenced by nodes within min(n+1, D) hops;	$\varphi(v_i) = \exp-(\frac{l}{\sigma})^2 * \sum_{l\in\min(n+1,D)} n_l(v_i)+1$
$\sigma\geq\frac{\sqrt{2}D}{3}$	influence within D hops, equal to closeness of nodes;	$\varphi(v_i) = 1 + \sum_{l\in D} n_l(v_i) * \exp-(\frac{l}{\sigma})^2$

4 Experiment

We evaluated our topology potential on the US politics books network compiled by V. Krebs. Nodes represent US politics books sold by the online bookseller Amazon. Edges represent frequent co-purchasing of books by the same buyers [7]. It is an undirected and unweighted network, which has 105 nodes and 441 edges. Average distance is 3.078755 and diameter of the whole network is 7.

We apply topology potential to measure the differentiation of each node in the network. During this experiment each node is supposed to be equal in mass, indicating the same influence. First we analyze how different influence region to impact on the same network, where regions vary from the optimal, 1 hop to 7 hop. Table 2 shows the top 8 largest nodes of different regions. Here the optimal sigma is 0.942463.

Table 3 shows the potential values of optimal, 3 hops and 7 hops, selected from the above ranges randomly. Using different influence regions to evaluate position differentiation of each node, we can see that wider of influence region, larger of node potential value.

Table 2. Top 8 largest nodes of different regions in US politics network

Optimal	1 hop	2 hop	3 hop	4 hop	5 hop	6 hop	7 hop
13	13	13	13	31	31	31	31
9	9	9	9	13	10	10	59
85	85	4	4	9	13	59	10
4	4	85	31	4	9	50	50
73	73	73	73	73	73	8	8
67	67	67	85	67	4	73	15
74	74	74	67	10	59	15	77
31	31	31	74	85	67	13	73

Table 3. Potential values of optimal, 3 hops and 7 hops in US politics network

O-Node	O-Value	3-Node	3-Value	5-Node	5-Value	7-Node	7-Value
13	9.109617	13	20.401069	31	43.727529	31	62.478852
9	9.109617	9	20.265734	10	41.552984	59	61.395356
85	8.460847	4	19.436460	13	41.142943	10	61.233185
4	8.460847	31	19.172210	9	40.712326	50	60.996966
73	8.136463	73	18.590559	73	40.533031	8	60.806942
67	7.812078	85	18.589094	4	40.027302	15	59.634202
74	7.812078	67	18.527396	59	39.835405	77	59.510396
31	7.487693	74	17.613376	67	39.710532	73	59.235671

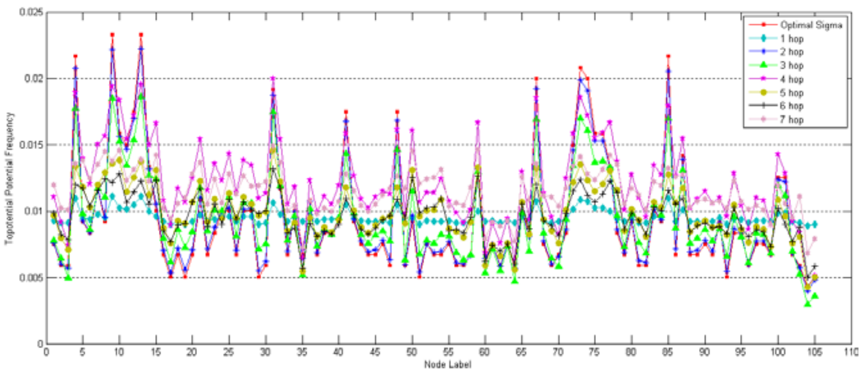

Fig. 1. Comparison among values of node topology potential with different influence regions on US politics network. Note that the optimal σ is better than any other factors because it can reflect the largest differentiation of the node position in the network. As depict with the red square of solid.

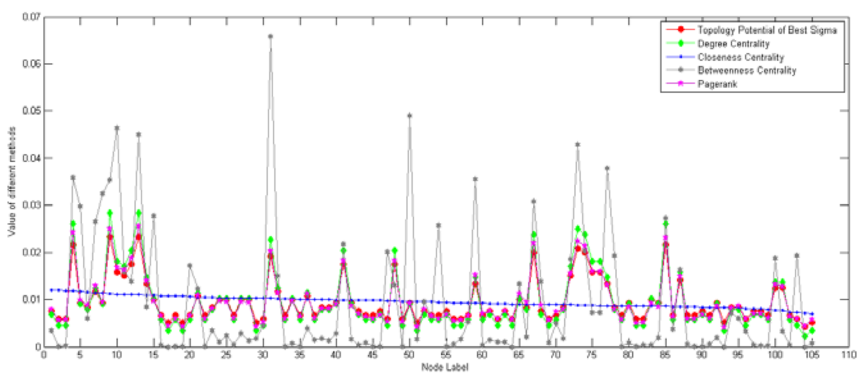

Fig. 2. Comparison among node values of different measurements on US politics network. Note that trend of optimal sigma is close to the degree and PageRank. While the trend of closeness with blue is relatively flat. It is hard to reflect differentiation of node position. There are also many nodes in betweenness with values of 0.

In order to compare the trend of different influence regions impact on the same network. We deal with normalization to node potential value, illustrated in Figure 1.

To validate how range of influence factor impact on other traditional measurements, as discussed in part 3.3. We compare all the results on the US politics network depicted in Figure 2. Here we adopt the optimal σ (0.942463). Evidences prove the topology potential is approximately equivalent to results of degree and other popular method PageRank. However closeness of node is hard to reflect the differentiation of node position in topology. And a lot of nodes' betweenness equal 0. That is not reasonable for differentiate nodes importance.

5 Conclusion

Here we introduce topology potential to evaluate interaction ability of node and locality of complex networks structure. It is very flexible by adjusting influence factor and inherently reflect traditional measurements. The presented measurement here lays foundations for future theoretical studies on networked data mining. This work has been supported by Program No. 2007CB310803. Future studies will be focus on community member identification, community discovery, backbone reduction, etc.

References

1. Katy, B., Soma, S., Alessandro, V.: Network Science. Annual Review of Information Science and Technology 41, 537–607 (2007)
2. Newman, M.E.J.: The Structure and Function of Complex Networks. SIAM Rev. 45, 167–256 (2003)
3. Watts, D.J., Strogatz, S.H.: Collective Dynamics of 'small-world' Networks. Nature 393, 440–442 (1998)
4. Newman, M.E.J.: Assortative Mixing in Networks. Phys. Rev. Lett. 89, 208701 (2002)
5. Colizza, V., Flammini, A., Serrano, M.A., Vespignani, A.: Detecting Rich-Club Ordering in Complex Networks. Nature Phys. 2, 110–115 (2006)
6. Strogatz, S.H.: Exploring Complex Networks. Nature 410, 268–276 (2001)
7. http://www-personal.umich.edu/~mejn/netdata/