Analysis and Calculation of Odd Cycle Centrality Aimed for Vertex-Cover

Ting Wang
LMIB and School of Mathematics and Systems Science
Beihang University
Beijing, China
wangting1232@163.com

Wei Wei
LMIB and School of Mathematics and Systems Science
Beihang University
Beijing, China
weiw@buaa.edu.cn

***Abstract*—We propose a kind of centrality, the odd cycle centrality which has strong relationship with solving minimum vertex-cover. The formal definition of the odd cycle centrality with its comparison to other centralities is given. Besides, as a variant of MAX-2-XORSAT, it is analyzed in the viewpoint of solving linear equations set. To calculate this centrality and view its effect on different graph structures, a random-walk based method is built to realize the centrality of each node on random graphs and scale-free networks. Using the importance order such as centrality provides one way to understand the topological structure or functions of a network, and it also offers a new viewpoint to solve the constraint satisfaction problems and recognize their complexity.**

Keywords—centrality; minimum vertex cover; MAX-2-XORSAT; random walk

I. INTRODUCTION

Minimum vertex cover problem is one of the six basic NP-complete problems [1], it can't get the optimal solution in polynomial time, unless P = NP, and is acknowledged as the classic problems of theoretical computer science. After Krap proved the NP completeness of minimum vertex cover problem [2], it becomes a hot topic attracting many attentions from studies, such as mathematicians, physicists and computer scientists. Research on the problem not only has important theoretical significance since it could be converted to minimum independent set problem and group problem [3], but also it has been widely used in immunization strategies in networks [4] and monitoring of internet traffic [5]. The study of the problem comes in several varieties: improving the efficiency of existing algorithms, studying the algorithms of particular graph, research for the restrictive variant problem and the solution space [6]. For the algorithms of the problem, there are heuristic algorithms, greedy algorithms, DNA genetic algorithms, fixed parameter algorithms, and so on.

As introduced in [7], nodes on a graph for vertex-cover have different status which performs as the importance of different nodes, such as backbone nodes and long-range frustration nodes [8]. But as the NP-completeness of vertex-cover, to identify the nodes' importance is not an easy job. How to probe the importance of the node in a complex network environment have become a basic problem of complex network research, and a fundamental problem in the field of graph theory [9]. With the complex network research, especially a lot of the actual network abstraction of complex systems in recent years, it is shown that different structural and statistical characteristics are studied for different aims. Important nodes in complex networks refer to some special nodes that affect the network structure and function in a larger extent, when comparing to other nodes in networks. Over the years, network researchers have introduced a large number of centrality indices [10], which measure the varying importance of the nodes in a network according to one criterion or many. These indices has been proven to be a valuable research in the social network by actors, as well as in other networks, including the citation network, computer network and biological network [11].

In [6, 7], the odd cycles on a graph perform great difficulties on finding the optimal solutions and solution space of minimum vertex-cover, and how to delete the fewest nodes to break all the odd cycles on a graph is the kernel problem of recognizing the complexity of minimum vertex-cover. This fact inspires us to detect the most important nodes which involve in the oddest cycles. In this paper, we will define a new centrality - *odd cycle centrality*, provide some strict formulation of it and build a random-walk method [12] to calculate it. And, the odd cycle centrality is calculated on two kinds of networks - the random graph and the scale-free network.

II. DEFINITION OF THE ODD CYCLE CENTRALITY

A vertex-cover of an undirected graph $G= (V, E)$ is a subset V' of V such that if edge (u, v) is an edge of G, then either u in V' or v in V' (or both). The minimum vertex cover of G is to find the minimum subset V' can cover G.

The existence of odd cycle is the biggest difficulty in minimum vertex-cover problem, especially when there is no leaf-removal core [7, 13]. The odd cycle is a cycle with odd nodes, and for minimum vertex cover, in an odd cycle with $2k+1$ nodes there should be $k+1$ covered nodes, the distribution of which can be of many choices [7]. Using the relationship between maximum matching and vertex cover [6], the old cycles make the maximum matching group in the solution diagram are not enough to fulfill the coverage.

Sponsored by the Natural Science Foundation of China (Grant No. 11201019) and the Fundamental Research Funds for the Central Universities.

Complex odd cycle structure makes algorithm face great obstacles to choose covered nodes efficiently and produces high complexity. On an odd cycle, we can choose any node covered and achieve coverage of the rest using the Konig theory [14] (*odd cycle breaking* operation), but unsuitable choice of the covered node in the odd cycle breaking operation after several such iterations may cause the increase of energy unnecessarily, so as to make the solution diagram collapse to the subspace of solutions. For many odd cycles couple together, we can choose some certain nodes to cover with the rest part of bipartite graph, and suitable choice of covered nodes to break all the odd cycles will lead to minimum vertex covers. But in order to get the exact solution/solution diagram in the end, we hope that with the least amount of such chosen covered nodes to break the odd cycles, by which the minimum vertex cover problem can be transformed into MAX-CUT problem [15]. By the conclusion of the MAX-CUT, after removing the edges not belonging to the MAX-CUT, we get compatible cycle. However, MAX-CUT is another one of the NP -complete problems and still can't be easily solved.

In Fig. 1, there are three odd cycles (2-3-4, 2-4-5 and 6-7-8), the black nodes (node 4 and 7) are chosen to be covered and break all the odd cycles, the rest is a bipartite graph ({2, 6}∈M and {1,3,5,8}∈N).

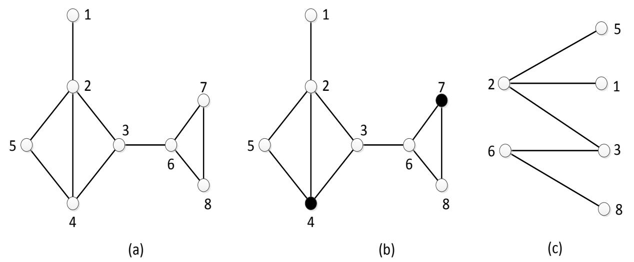

Fig. 1. A graph (a) with 8 nodes and 10 edges, there are 3 odd cycles in the graph, and after deleting the set of black nodes (b), the rest is a bipartite graph (c).

Since the importance of odd cycles, the *odd cycle centrality* of a node *v* is given by the expression:

$$C_o(v) = \frac{\sigma_o(v)}{\sigma(v)},$$

where $C_o(v)$ is the odd cycle centrality of node v, $\sigma_o(v)$ is the total number of odd cycles containing node v, $\sigma(v)$ is the total number of cycles containing node v. Note that the odd cycle centrality of a node scales with the number of cycles of the node as implied by the denominator, so that $C_o(v) \in [0, 1]$.

III. Comparison with Other Centrality

In order to compare with odd cycle centrality we defined, we choose four kinds of the most commonly used centralities (Degree centrality, Closeness centrality, Betweeness centrality and Eigenvector centrality). The four central principles are not completely the same, and except the degree centrality using node relevance principle, the rest of the three kinds of centralities are adopting the principle of the shortest path. Frist, we will make brief introduction of four centralities, and then give one simple example to calculate all these five centralities.

Degree centrality [16] refers to the degree of a node, i.e., the number of edges connected to one node. Degree centrality is the most simple and intuitive way to examine the importance of nodes, which is based on the node nearest neighbors. But the degree centrality only considers partial information nodes, not depth discussion of surrounding environment, thus it is not accurate enough in many cases.

Closeness centrality [17] was defined by Bavelas as the reciprocal of the sum of shortest distance from one node to others on the graph. From the perspective of the shortest path, closeness centrality can be regarded as a measure of how long it will take information to spread from a given node to others in the network. Larger closeness centrality means the more center position of the node, and vice versa.

Betweeness centrality [18] is defined as the ratio of the shortest paths containing some node *v* and all shortest paths between all the nodes' pairs of the graph. Betweenness centrality portrays the control to network flow of node along the shortest path, and the more number of the shortest path through a node, the more important of this node will be. But the high time complexity makes betweeness centrality limited in practical applications.

Eigenvector centrality [19] is defined as the largest eigenvector of the adjacency matrix eigenvalues correspondingly (we can calculate the normalized vector). Eigenvector centrality considers not only the importance of the number of its neighbor nodes (that is, the node degree), but also the importance of each neighbor node on the graph.

As an example, comparison of the five kinds of centralities in Fig. 1 (a) is shown in the following table.

TABLE I. The five centralities of Fig. 1. (a)

Nodes	Centrality				
	Odd Cycle	***Degree***	***Closeness***	***Betweeness***	***Eigenvector***
Node 1	0	1	1/18	7/32	0.0732[a]
Node 2	2/3	4	1/12	17/32	0.2051
Node 3	1/2	2	1/17	11/32	0.1395
Node 4	2/3	3	1/13	11/32	0.1859
Node 5	1/2	3	1/11	23/32	0.1766
Node 6	1	3	1/13	20/32	0.1042
Node 7	1	2	1/18	8/32	0.0578
Node 8	1	2	1/18	8/32	0.0578

a. The normalized vector.

IV. The correlation with 2-XORSAT problem

SAT problem is the first problem proved NP-complete in the history and the XORSAT problem is a variant of SAT. For 2-XOR satisfiability problem (2-XORSAT) [20], each instance is a formula with a conjunction of Boolean equations in the form $x+y=0$ or $x+y=1$. MAX-2-XORSAT [21] is an optimization problem, which asks for the maximum number of Boolean equations can be satisfied by any assignment of the variables in a 2-XORSAT formula. MAX-2-XORSAT and minimum vertex-cover problem are closely linked, and the odd cycle centrality can be transformed to some variant of MAX-2-XORSAT. For a given graph to calculate its odd cycle centrality, let each node i having a Boolean variable $x_i \in \{0, 1\}$, if x_i connects with x_j, than we get equation $x_i + x_j = 1$. If the graph has odd cycle, then the corresponding equations set have conflict equations and no solution. The sum of equations on one odd cycle is 1 and the sum on an even cycle is 0. If the equations for the whole graph have no solution satisfying all, it means the existence of odd cycle on the graph. The MAX-2-XORSAT problem can be converted to delete some equations (edges) and make it has solution. For the calculation of odd cycle centrality, what we aim is to delete some variables and make it has solution. MAX-2-XORSAT is to solve the values of $(x_1, x_2, \ldots, x_n)$ that can satisfy the most equations, i.e., deleting the least number of equations to make it has solution. In this paper we consider such nodes having higher odd cycle centrality, the deletion of which breaks more odd cycles.

If we delete node x_i (it means covered node x_i in the vertex-cover problem), all the edges connected to the node x_i have been deleted. For the equations, all equations associated with x_i are also removed. However, if we delete one edge of the graph, we only need to remove this corresponding equation.

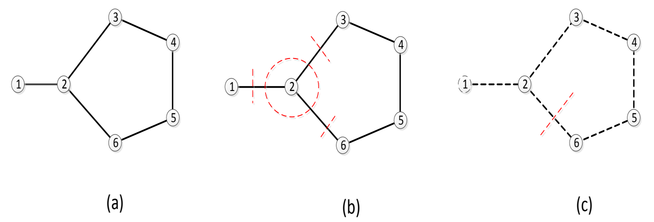

Fig. 2. A graph with 6 nodes and 6 edges, there is an odd cycle in the graph (2-3-4-5-6-2). (a) is the original graph, (b) is the graph we'd like to delete node 2, (c) is the graph we'd like to delete edge 2-6. The red dotted line is the delete tag.

For example, Fig.2 (a) is the original graph, (b) is the graph which deletes node x_2, and (c) is the graph which removes the edge connecting x_2 and x_6. Fig. 2 (a-c) can be converted to the corresponding equations set (a-c) as:

$$\begin{cases} x_1 + x_2 = 1 \\ x_2 + x_3 = 1 \\ x_2 + x_6 = 1 \\ x_3 + x_4 = 1 \\ x_4 + x_5 = 1 \\ x_5 + x_6 = 1 \end{cases} \text{(a)}, \quad \begin{cases} x_3 + x_4 = 1 \\ x_4 + x_5 = 1 \\ x_5 + x_6 = 1 \end{cases} \text{(b)}, \quad \begin{cases} x_1 + x_2 = 1 \\ x_2 + x_3 = 1 \\ x_3 + x_4 = 1 \\ x_4 + x_5 = 1 \\ x_5 + x_6 = 1 \end{cases} \text{(c)},$$

where (a) is the primitive equation set which corresponds to the full graph. The equations set (b) removes the first, second and third equations of set (a) which all constrain x_2. The equations set (c) removes the third equation of set (a) which contains x_2 and x_6.

V. Calculation of Odd Cycle Centrality

As the analysis in the above section, directly calculating the odd cycle centrality is very hard, and in this chapter we will give method to calculate the odd cycle centrality of the graph by numerical experiments. Besides, we will compare the odd cycle centrality difference between the random graph and the scale-free network.

In mathematics, random graph [22] is the general term to refer to generating graphs over some probability distributions. A random graph is obtained by starting with a set of n-isolated vertices and adding successive edges between them at random. The average degree of is given by $c = m / n$, where m is the number of edges and n is the number of nodes in the graph. When generating a graph of average degree c, we first set n nodes, and then randomly generate cn edges among the nodes.

The Barabási–Albert (BA) model [23] is an algorithmic model for generating random scale-free networks. BA model has two features, one is the growth, which refers to the size of network constantly increasing in the research of network; the other is the preferential attachment mechanism, which means that the new network node has more tend to be connected with the nodes having larger degrees. A scale-free network [24] is a network whose degree distribution follows a power law, at least asymptotically. That is, $P(k)$, the fraction of nodes having k connections to other nodes in the network goes for large values of k as $P(k) \sim k^{-r}$ where r is a parameter whose value is typically in the range $2 \leq r \leq 3$. Scale-free network widely exists in the information exchange network, social networks and biological networks [25].

Based on the strategy of random walk, we could calculate the odd cycle centrality of each node of a given graph. Through the following steps, we can get the odd circle centrality of node i.

Step1: mark the state of node i to +1;

Step2: Randomly select the neighbor node j of node i, the state of node j is -1, and the current node is j.

Step3: Consider the current node is l, randomly choose a neighbor l' of it, and the state of l' is the opposite number of that of l. This is called a *jump*, and for every jump, the node state is multiplied by -1.

Step4: After some jumps, there is a chance to jump back to the original node i. When it jumps back to node i, we will record recycle times by increasing 1 from R to R+1. When it jumps back to node i and the state of node i changes to -1, we will record state changes by increasing 1 from S to S+1.

Step5: Finally, define the ratio of state change times and recycle times (S/R) as the odd cycle centrality of node i.

By the above steps, we can obtain an approximated odd cycle centrality for each node i after a large number of jumps. As the numerical results in Fig.3, for the random graphs, the odd cycle centrality increases as the increase of average degree which reflects the more difficulty in solving Vertex-Cover instances. But for the scale-free graph by BA model, there is quite different effect on the distribution of odd cycle centrality: about one third nodes have very low odd cycle centrality and the rest nodes distribute relatively evenly in a large interval, which implicates that solving Vertex-Cover instances on scale-free graphs should be less complicated than on random graphs.

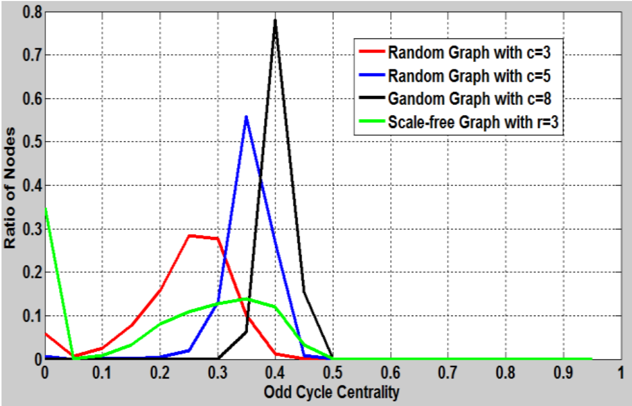

Fig. 3. Numerical results of the odd cycle centrality for random graphs with average degrees c=3,5,8 and scale-free graphs (BA model) with degree distribution $P(k) \sim k^{-r}$ (r=3). All the experiments are performed for 1000 instances with node number n=500 and 10^6 jumps.

The achieving of the odd cycle centrality can help us solve the minimum vertex-covers. For the nodes with higher odd cycle centrality, there is more possibility that it is a covered backbone in the solution space of minimum vertex-covers; and after breaking the odd cycles with lowest cost, finding the minimum vertex-covers is relatively easy under a bipartite graph [6]. Though the proposed random-walk method may not be quite strict, it should be an asymptotic one as the jumps number goes to infinity.

VI. Conclusion

A kind of centrality named odd cycle centrality is proposed to identify the node which is involved in the most number of odd cycles of a graph. This centrality has direct correspondence with solving the minimum vertex-covers and can be obtained by a random-walk method. Indeed, the centralities provide a new viewpoint to detect the node's importance in constraint satisfaction problems. By properly assigning an importance order, the nodes having high centrality have high priority to be considered to satisfy the constraints.

Acknowledgment

This work is supported by the Fundamental Research Funds for the Central Universities, the National Natural Science Foundation of China (Grant No.11201019).

References

[1] Cook S A, The complexity of theorem-proving procedures, ACM, 1971: 151-158.

[2] Edmonds J, Karp R M, "Theoretical improvements in algorithmic efficiency for network flow problems," J. Journal of the ACM (JACM), 1972, 19(2), pp. 248-264.

[3] Hochbaum D S, Approximating covering and packing problems: set cover, vertex cover, independent set, and related problems //Approximation algorithms for NP-hard problems. PWS Publishing Co., 1996, pp. 94-143.

[4] Gómez-Gardenes J, Echenique P, Moreno Y, "Immunization of real complex communication networks," J. The European Physical Journal B-Condensed Matter and Complex Systems, 2006, 49(2), pp. 259-264.

[5] Breitbart, Y., Chan, C. Y., Garofalakis, M., Rastogi, R., and Silberschatz, A, Efficiently monitoring bandwidth and latency in IP networks//INFOCOM 2001. Twentieth Annual Joint Conference of the IEEE Computer and Communications Societies. Proceedings. IEEE. IEEE, 2001, 2, pp. 933-942.

[6] Wei, W., Zhang, Y., Wang, T., Li, B., Niu, B., and Zheng, Z, "Research on Solution Space of Bipartite Graph Vertex-Cover by Maximum Matchings," J. arXiv preprint arXiv:1505.06955, 2015.

[7] Wei, W., Zhang, R., Guo, B., and Zheng, Z, "Detecting the solution space of vertex cover by mutual determinations and backbones," J. Physical Review E, 2012, 86(1): 016112.

[8] Zhou H "Long-range frustration in a spin-glass model of the vertex-cover problem," J. Physical review letters, 2005, 94(21): 217203.

[9] Strogatz S H, "Exploring complex networks," J. Nature, 2001, 410(6825), pp. 268-276.

[10] Friedkin N E, "Theoretical foundations for centrality measures," J. American journal of Sociology, 1991, pp. 1478-1504.

[11] Freeman L C, "Centrality in social networks conceptual clarification" J. Social networks, 1979, 1(3), pp. 215-239.

[12] Newman M E J, "A measure of betweenness centrality based on random walks," J. Social networks, 2005, 27(1), pp. 39-54.

[13] Bauer M, Golinelli O, "Core percolation in random graphs: a critical phenomena analysis," J. The European Physical Journal B-Condensed Matter and Complex Systems, 2001, 24(3), pp. 339-352.

[14] Bondy J A, Murty U S R, Graph theory with applications, London: Macmillan, 1976.

[15] Festa, P., Pardalos, P. M., Resende, M. G., and Ribeiro, C. C, "Randomized heuristics for the MAX-CUT problem," J. Optimization methods and software, 2002, 17(6), pp. 1033-1058.

[16] Opsahl T, Agneessens F, Skvoretz J, "Node centrality in weighted networks: Generalizing degree and shortest paths," J. Social Networks, 2010, 32(3), pp. 245-251.

[17] Okamoto K, Chen W, Li X Y, Ranking of closeness centrality for large-scale social networks, //Frontiers in Algorithmics. Springer Berlin Heidelberg, 2008, pp.186-195.

[18] Newman M E J, "A measure of betweenness centrality based on random walks," J. Social networks, 2005, 27(1), pp. 39-54.

[19] Bonacich P, "Some unique properties of eigenvector centrality," J. Social Networks, 2007, 29(4), pp. 555-564.

[20] Hervé D, Vlady R, "Random 2 XORSAT Phase Transition," J. Algorithmica, 2011, 59(1), pp. 48-65.

[21] Rasendrahasina V, Ravelomanana V, Limit theorems for random MAX-2-XORSAT, //LATIN 2010: Theoretical Informatics. Springer Berlin Heidelberg, 2010, pp. 320-331.

[22] Bollobás B, Random graphs, Springer New York, 1998.

[23] Barabási A L, Albert R, "Emergence of scaling in random networks," J. Science, 1999, 286(5439), pp. 509-512.

[24] Pastor-Satorras R, Vespignani A, "Epidemic spreading in scale-free networks," J. Physical review letters, 2001, 86(14), pp. 3200.

[25] Wang W X, Wang B H, Yin C Y, and Zhou, T, "Traffic dynamics based on local routing protocol on a scale-free network," J. Physical Review E, 2006, 73(2): 026111.