Available Online at www.ijcsmc.com

International Journal of Computer Science and Mobile Computing

A Monthly Journal of Computer Science and Information Technology

ISSN 2320–088X
IMPACT FACTOR: 5.258

IJCSMC, Vol. 5, Issue. 5, May 2016, pg.281 – 287

Portrayal Analogy of BFS and MHRW for Online Social Network Expository

Gayathri.G.S[1], Bhagyashree.P[2]
[1]Assistant Professor, Department of Computer Science and Engineering, India
[2]M.Tech Student, Department of Computer Science and Engineering, India
New Horizon College of Engineering, Bangalore, India
[1] gayathrigummaraj@gmail.com; [2] bhagyashree.pandian@gmail.com

Abstract --- Several different communities experience different opinions regarding the Online Social Network [OSN]. The examination of these OSN and its knowledge is called as Social Network Analysis (SNA). Having the capacity to keep the graph scale little while catching the properties of the imaginative social diagram, graph inspecting gives an effective, nevertheless economical answer for social organize examination. The test is the way to make a little, yet illustrative test out of the monstrous social graph with millions then again even billions of nodes. A few testing calculations have been proposed in the past concentrates, however these need reasonable assessment and examination among them. In this system, we examine the state of art diagram examining calculations and assess their execution on some generally perceived diagram properties on coordinated graphs utilizing vast scale interpersonal organization datasets. We assess not just the normally utilized degree dispersion, additionally grouping coefficient, which measures how very much associated are the neighbors of a node in a diagram. In this paper, we have shown the portrayal graphs based on various parameters for BFS and MHRW algorithms.

Keywords: Social Network, SNA, Graph Sampling, BFS, MHRW.

I. INTRODUCTION

The most recent couple of years have seen a dangerous development of online social network [OSN] that have pulled in most consideration from everywhere throughout the world. Facebook, an interpersonal organization benefit, has pulled in more than 600 million dynamic clients as of January, 2011. Twitter, a social micro-blogging benefit known as the "SMS of the Internet", has more than 190 million clients who produce more than 65 million "tweets" each day. The tremendous client base of these OSNs gives an open stage for informal community investigation including client conduct estimations, social cooperation portrayal, and data proliferation thinks about. In any case, the gigantic size of informal community graphs obstructs specialists from a superior comprehension of these diagrams.

On one hand, it is difficult to secure the finish diagram of a system. While arrange overseers are unwilling to give their information to specialists, slithering the finish diagram of these informal communities is constantly unthinkable, particularly considering the get to guidelines set by the systems and the measure of time it would take. Then again, even with right now accessible datasets of these systems, preparing them requires costly and very much prepared PC bunches, and additionally comprehensive time and calculation overhead. Then again, graph inspecting gives an effective, yet reasonable arrangement. By selecting a delegate subset of the first graph, diagram sampling can make the graph scale small while keeping the characteristics of the original social graph.

The following key points describe the methodology to collect the online social network data.
They are,
(i) Complete dataset is unavailable.
(ii) OSNs unwilling to share company's data.
(iii) Large size of social network graph.
(iv) Access limitations.

The following algorithms illustrate the nature of proposed work and its associated ruling methodologies.
They are,

(a) Breadth First Search (BFS)
(b) Metropolis Hasting Random Walk (MHRW)

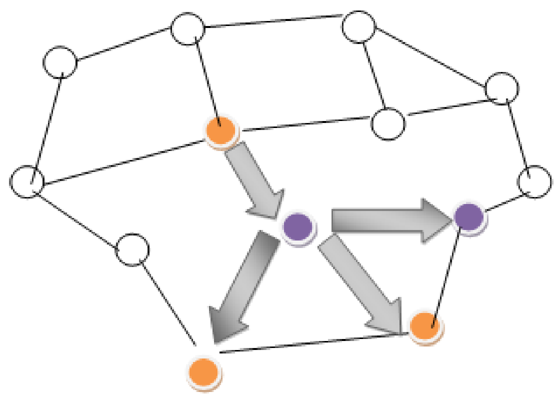

Fig.1. Sampling Methodology

II. **RELATED WORK ANALYSIS**

A few examining calculations have been proposed for diagram testing. Breadth-First Sampling (BFS) and Random-Walk [RW] are the most surely understood examining calculations and have been utilized as a part of numerous territories. In any case, past studies demonstrate that BFS and RW make tests one-sided to high-degree s. Metropolis-Hasting Random Walk (MHRW) is utilized to get fair-minded specimens in undirected social charts, i.e., keeping the degree appropriation of the first diagram unaltered. USDSG makes MHRW material in coordinated social charts by considering all the unidirectional edges as bidirectional edges.

MHRW, USDSG, and FS all contrast their calculation and RW on node degree appropriation utilizing diverse datasets, while the face to face examinations between these recently proposed calculations is an empty. Additionally, other than node degree dissemination, other chart properties, for example, grouping coefficient have not been talked about compressively in the current studies, which constrains the extent of potential applications.

III. **CONTRIBUTIONS**

In the proposed graph sampling methodology, we attempt to investigate how existing calculations perform in keeping up various essential properties of unique social diagrams. The datasets we pick are all certifiable social diagrams and have been generally perceived in numerous other looks into. We assess these calculations considering Normalized Mean

Square Error (NMSE), Cumulative Distributive Function (CDF), Time, and Distance. As per our estimation consider, we find that these calculations perform differently on keeping up various diagram properties. Additionally, the execution is much corresponded with particular dataset. A calculation can carry on ineffectively in some datasets even in spite of the fact that it performs entirely well in another. We attempt to get a few bits of knowledge of this execution distinction by considering the graph properties. Portrayal of BFS and MHRW based on these four parameters are clearly observed. We find MHRW perform better in firmly associated graphs.

IV. PROPERTIES OF PARAMETERS CONSIDERED

Normalized Mean Square Error (NMSE): NMSE is an estimator of overall deviations between predicted and measured values. In the NMSE the deviations are summed instead of the differences. For this, reason the NMSE generally shows the most striking differences among models. If a model has a very low NMSE, then it is well performing both in space and time. On the other hand, high NMSE values indicate that the performance is poor.

Cumulative Distributive Function (CDF): CDF of a real-valued random variable X, or just distribution function of X, evaluated as x, is the probability that X will take a value less than or equal to x.

Time: Time is the indefinite continued progression of existence and events that occur in apparently irreversible succession from the past through the present to the future. Time is a component quantity of various measurements used to sequence events, to compare the duration of events or the intervals between them.

Distance: Distance is a numerical description of far apart objects. That is, the interval between two points of time.

V. BREADTH FIRST SEARCH (BFS)

Breadth First Search is a testing calculation which has been generally concentrated on and connected in client conduct investigation of OSNs, estimation and topological attributes examination of OSNs. BFS can discover the nodes nearest to the underlying node and it is utilized for exploring and examination. BFS works in the accompanying way. It begins from a haphazardly chosen seed. There are two lines in the testing handle: line Sampled stores inspected nodes, while line Processed stores nodes that have been prepared. By "prepared" we mean tested or with one of their neighbors inspected. At first, the seed is put away in line Processed. At every circle, the primary ‘v’ in line Processed is moved to line Sampled, and every one of the neighbors of ‘v’ are embedded into line Processed, unless the node has as of now been prepared, i.e., in line Processed. The procedure circles until the altered spending plan is come to. It is conceivable that the financial backing is never come to if the underlying situates in a little confined sub-graph.

For this situation, another arbitrarily chosen seed is embedded to line Processed. Since the nodes in OSNs are generally exceedingly associated with each other, the risk of falling into this case is uncommon. It's realized that BFS is predisposition to high degree nodes, as is called attention to. In BFS, nodes with a higher degree will be gone by all the more as often as possible. This wonder will be appeared and we additionally demonstrate BFS gets higher nearby grouping coefficient than the first ones because of the predisposition.

The following description clearly illustrates the working of Breadth First Search algorithm.

Algorithm: Breadth First Search [BFS]

Step 1: Start at node v.
Step 2: Sampled- node v is visited.
Step 3: Nodes adjacent to Sampled node (v).
Step 4: Processed-adjacent nodes not visited.
Step 5: Pull out first from Processed and traverse to it.
Step 6: Go back to step 1.

VI. METROPOLIS HASTING RANDOM WALK

MHRW is a Markov Chain Monte Carlo [MCMC] calculation to acquire arbitrary tests as per the degree likelihood conveyance of the nodes. This is regularly hard to accomplish by specifically examining. In MHRW, a proposition capacity is planned in light of the likelihood dissemination. By arbitrarily tolerating or rejecting the proposition, the proposition work changes the move probabilities, making the examples merge to the likelihood appropriation. In this paper, we utilize MHRW to inexact the uniform circulation since we need the nodes be gone by consistently. At first, a haphazardly chosen node with non-zero degree is set as the seed. We characterize the proposition work as $x^{(i)}$ = ki, which is the level of node v.

From v's neighbors, MHRW arbitrarily picks a w, and after that creates an irregular number u from uniform dispersion U(0; 1). On the off chance that u, $x^{(i)} = x^{(z)}$, the proposition is acknowledged and the inspecting process will travel to i; else, it stays at z. Take note of that on the off chance that it stays at node v, it doesn't spend a cost, following the node's profile has been downloaded as of now. The proposition work changes the move probabilities along these lines: if the level of ($x^{(z)}$) is little, in spite of the fact that z will have a little opportunity to be picked as the applicant, there will be a high likelihood that the proposition will be Acknowledged once it happens. In this way the proposition work redressed the predisposition towards high-degree nodes.

MHRW stops when the monetary allowance is come to. MHRW was initially intended for undirected diagrams. In a strategy called USDSG is produced in light of MHRW to work in coordinated diagrams. USDSG considers all the unidirectional edges as bidirectional edges. To apply USDSG, we have to change a guided chart Gd to a symmetric diagram G. Since this is the main distinction in the middle of MHRW and USDSG, to be straightforward, we will utilize term MHRW to speak to both the first MHRW and USDSG starting now and into the foreseeable future. MHRW considers all the copied nodes as legitimate nodes. These copied nodes make the conveyance merge to uniform circulation. We don't have to consider the situation when we stroll to a node with zero degree aside from the seed, subsequent to the way that a node can be gone by intrinsically requests that its degree is not 0.

The following description clearly illustrates the working of Metropolis Hasting Random Walk algorithm.

Algorithm: Metropolis Hasting Random Walk

Step 1: Start at $x^{(i)}$

Step 2: Randomly choose $x^{(z)}$
Step 3: Generate random number u from U(0,1)
Step 4: If $u \leq \frac{x^{(i)}}{x^{(z)}}$
Step 5: $x^{(i+1)} = x^{(z)}$
else
Step 6: $x^{(i+1)} = x^{(i)}$

For this system, we consider the dataset as Wikipedia and work based on that. The illustration sample of Wikipedia dataset is given below:

Site	YouTube	Flickr	LiveJournal	Orkut
Nodes(mill)	1.1	1.8	5.2	3
Links (mill)	5.9	22	72	223

VII. **EXPERIMENTAL RESULTS**

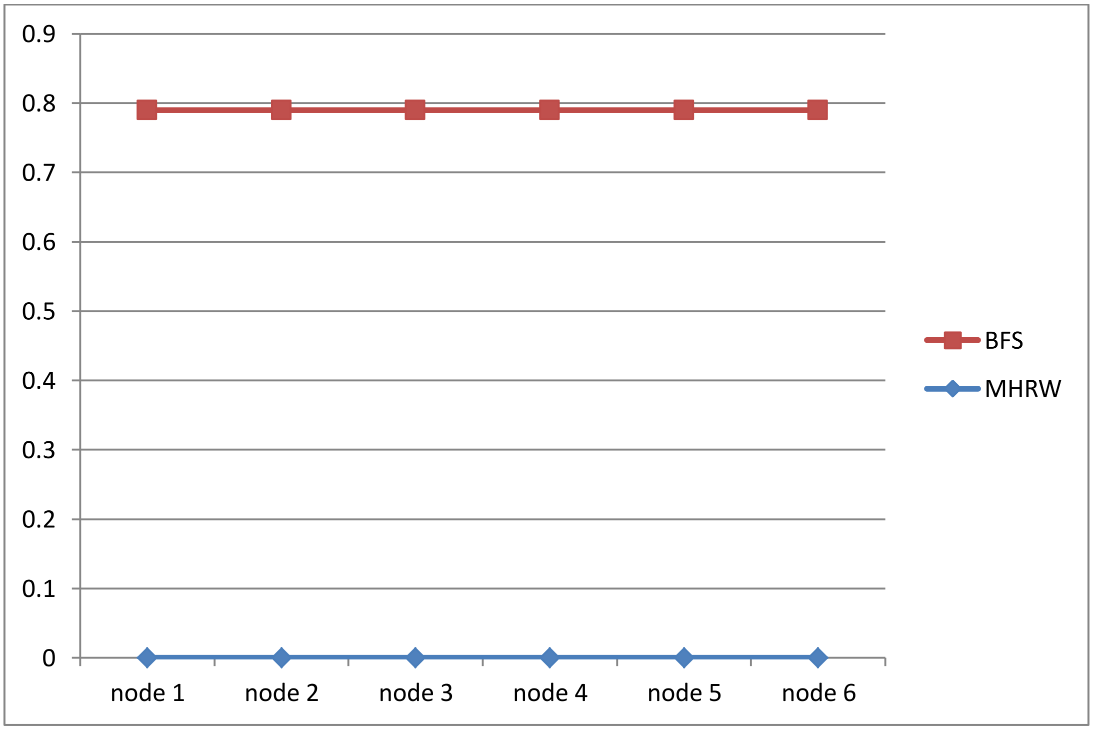

Fig: Depicts Nodes Vs NMSE

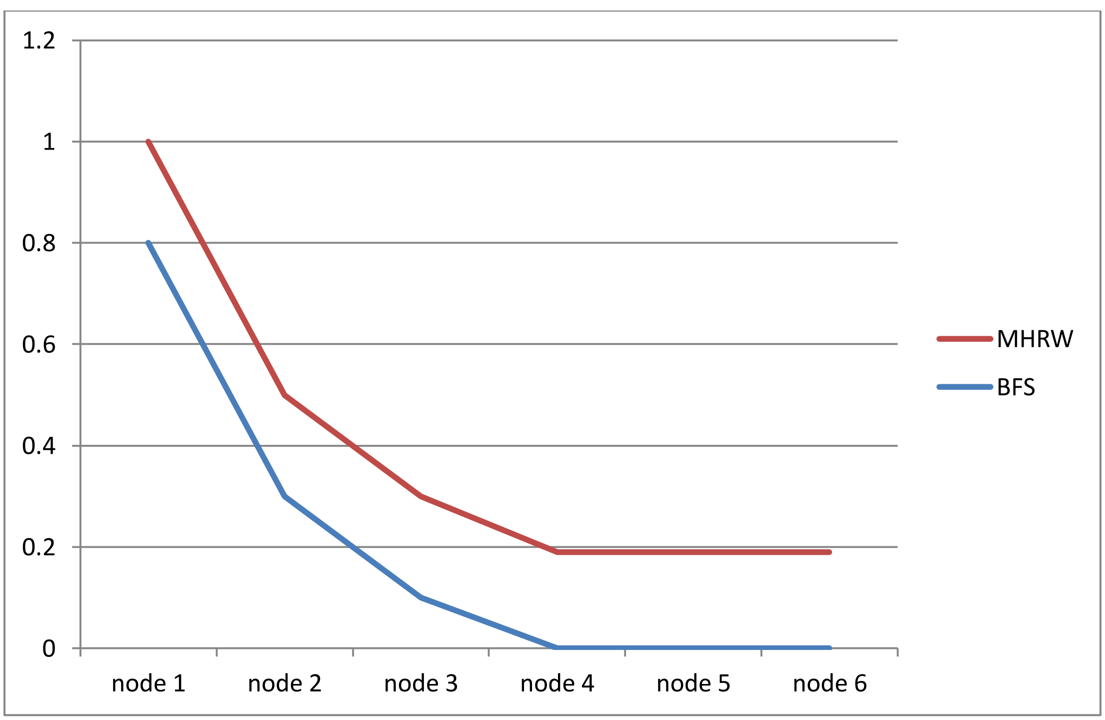

Fig: Depicts Nodes Vs CDF

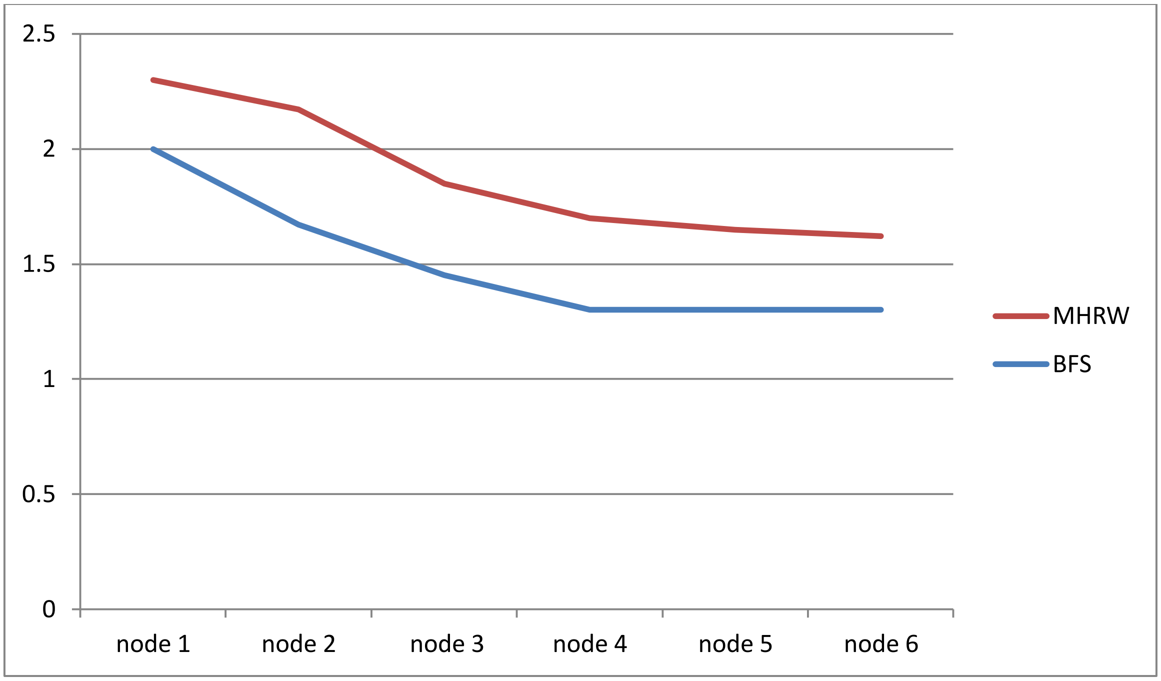

Fig: Depicts Node-cycle Vs Time

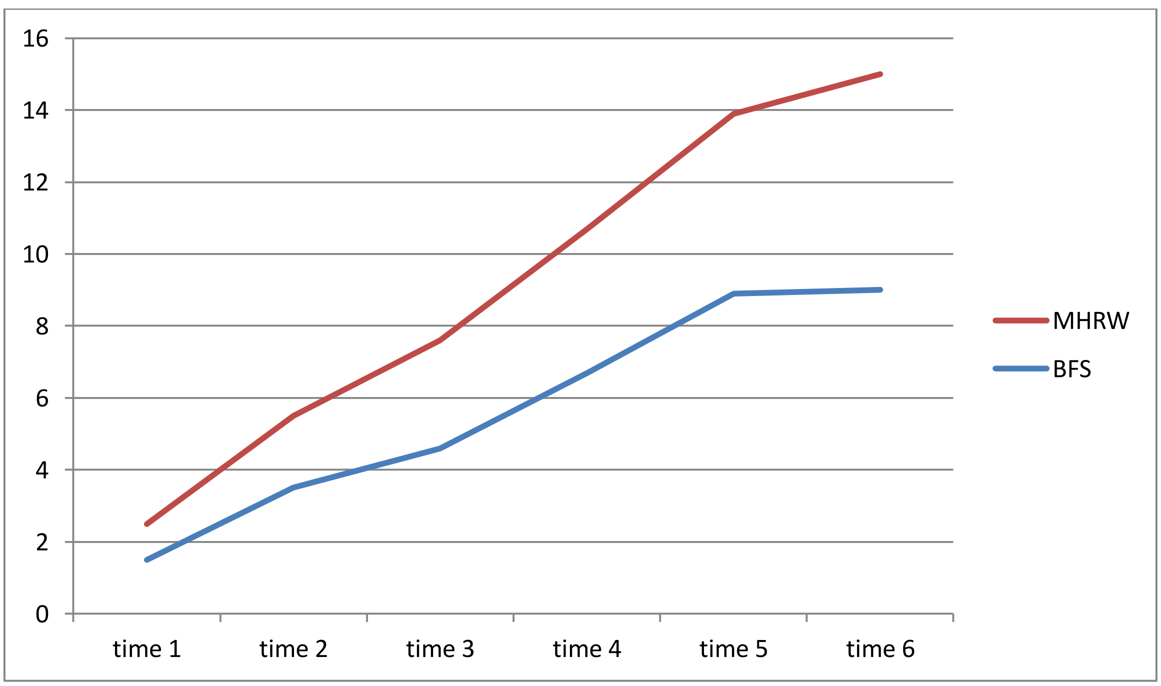

Fig: Depicts Time Vs Distance

VIII. CONCLUSION AND FUTURE WORK

In this framework we led an extensive study on a few examining strategies in social network graphs. We dissected how these examining techniques perform in keeping up the properties of the parameters graph models. Portrayal of BFS and MHRW is shown in the plots based on NMSE, CDF, Distance and Time. From these plots we can conclude that MHRW performs well than BFS with respect to the datasets considered. The future enhancement can be considering any other parameters and datasets.

REFERENCES

[1] MSNBC, “Goldman to clients: Facebook has 600 million users,” http://www.msnbc.msn.com/id/40929239/ns/technology-and-sciencetech-and-gadgets/.
[2] Quantcast.com, “Twitter.com - Quantcast Audience Profile,” http://www.quantcast.com/twitter.com.
[3] J. Leskovec and C. Faloutsos, “Sampling from Large Graphs,” In Proc. of ACM SIGKDD, 2006.
[4] C. Wilson, B. Boe, A. Sala, K. P. N. Puttaswamy, and B. Y. Zhao, “User Interactions in Social Networks and their Implications,” In Proc. of ACM EuroSys, 2009.
[5] M. Gjoka, M. Kurant, C. T Butts, and A. Markopoulou, “Walking in Facebook: A Case Study of Unbiased Sampling of OSNs,” In Proc. of IEEE INFOCOM, 2010.
[6] T. Wang, Y. Chen, Z. Zhang, P. Sun, B. Deng, and X. Li, “Unbiased Sampling in Directed Social Graph,” In ACM SIGCOMM Computer Communication Review, 40(4):401-402, 2010.
[7] B. Ribeiro and D. Towsley, “Estimating and Sampling Graphs with Multidimensional Random Walks,” In Proc. of ACM IMC, 2010.
[8] D. J. Watts and S. Strogatz, “Collective Dynamics of ‘Small-World’ Networks,” Nature 393(6684): 440-442, 1998.
[9] Minas Gjoka, “Measurement of Online Social Networks,” UC Irvine PhD Thesis, 2010.
[10] H. Kwak, C. Lee, H. Park, and S. Moon, “What is Twitter, a Social Network or a News Media?” In Proc. of WWW, 2010.
[11] F. Benevenuto, T. Rodrigues, M. Cha, and V. Almeida, “Characterizing User Behavior in Online Social Networks,” In Proc. of ACM IMC, 2009.
[12] A. Sala, L. Cao, C. Wilson, R. Zablit, H. Zheng, and B. Y. Zhao, “Measurement-calibrated Graph Models for Social Network Experiments,” In Proc. of WWW, 2010.
[13] Stanford Large Network Dataset Collection. http://snap.stanford.edu/data/index.html.
[14] M. Kurant, A. Markopoulou, P. Thiran, “On the Bias of Breadth First Search (BFS) and of Other Graph Sampling Techniques,” International Teletraffic Congress, 2010.
[15] S. H. Lee, P. -J. Kim, and H. Jeong, “Statistical Properties of Sampled Networks,” Physical Review E, 2006.
[16] A. Mislove, M. Marcon, K. P. Gummadi, P. Druschel, and B. Bhattacharjee, “Measurement and Analysis of Online Social Networks,” In Proc. of ACM IMC, 2007.
[17] Y. Ahn, S. Han, H. Kwak, S. Moon, and H. Jeong, “Analysis of Topological Characteristics of Huge Online Social Networking Services,” In Proc. of WWW, 2007.

