Sampling algorithms for weighted networks

Social Network Analysis and Mining
December 2016, 6:60

Alireza Rezvanian, Mohammad Reza Meybodi

Original Article
First Online: 13 August 2016
DOI (Digital Object Identifier): 10.1007/s13278-016-0371-8

Cite this article as:
Rezvanian, A. & Meybodi, M.R. Soc. Netw. Anal. Min. (2016) 6: 60.
doi:10.1007/s13278-016-0371-8

Abstract

Many of the real-world networks, such as complex social networks, are intrinsically weighted networks, and therefore, traditional network models, such as binary network models, will result in losing much of the information contained in the edge weights of the networks and is not very realistic. In this paper, we propose that when the network model is chosen to be a weighted network, then the network measures such as degree centrality, clustering coefficient and eigenvector centrality must be redefined and new network sampling algorithms must be designed to take the weights of the edges of the network into consideration. In this paper, first, some network measures for weighted networks are presented and then, six network sampling algorithms are proposed for sampling weighted networks. The evaluation is done through simulations on real and synthetic weighted networks in terms of relative error, skew divergence, Pearson's correlation coefficient and the Kolmogorov–Smirnov statistic. A number of experiments have been conducted to compare the sampling algorithms for weighted networks proposed in this paper with their counterparts for unweighted networks. The experiments show that existing sampling

algorithms for unweighted networks will not produce good results as used for sampling weighted networks when compared to the algorithms proposed in this paper.

Keywords

Complex networks Social networks Social network analysis Network sampling Weighted networks

References

Beebe NH (2002) Nelson HF Beebe's bibliographies page. In: Nelson HF(ed) Beebe's bibliographies page. http://www.math.utah.edu/~beebe/bibliographies.html (http://www.math.utah.edu/%7ebeebe/bibliographies.html)

Blagus N, Šubelj L, Weiss G, Bajec M (2015) Sampling promotes community structure in social and information networks. Phys A 432:206–215
CrossRef (http://dx.doi.org/10.1016/j.physa.2015.03.048)

Chi G, Thill J-C, Tong D et al (2016) Uncovering regional characteristics from mobile phone data: a network science approach. Pap Reg Sci. doi:10.1111/pirs.12149:1-19 (http://dx.doi.org/10.1111/pirs.12149%3A1-19)

Cordeiro M, Sarmento RP, Gama J (2016) Dynamic community detection in evolving networks using locality modularity optimization. Soc Netw Anal Min 6:15. doi:10.1007/s13278-016-0325-1 (http://dx.doi.org/10.1007/s13278-016-0325-1)
CrossRef (http://dx.doi.org/10.1007/s13278-016-0325-1)

Dall'Asta L, Barrat A, Barthélemy M, Vespignani A, (2006) Vulnerability of weighted networks. J Stat Mech: Theory Exp 2006:P04006

Dijkstra EW (1959) A note on two problems in connexion with graphs. Numer Math 1:269–271
CrossRef (http://dx.doi.org/10.1007/BF01386390)
MATH (http://www.emis.de/MATH-item?0092.16002)
MathSciNet (http://www.ams.org/mathscinet-getitem?mr=107609)

Erdos P, Rényi A (1960) On the evolution of random graphs. Publ Math Instit Hung Acad Sci 5:17–61
MATH (http://www.emis.de/MATH-item?0103.16301)
MathSciNet (http://www.ams.org/mathscinet-getitem?mr=125031)

Frank O (2011) Survey sampling in networks. In: The SAGE Handbook of Social Network Analysis. SAGE publications, pp 370–388

Friedman M (1940) A comparison of alternative tests of significance for the problem of m rankings. Ann Math Stat 11:86–92
CrossRef (http://dx.doi.org/10.1214/aoms/1177731944)
MATH (http://www.emis.de/MATH-item?66.1305.08)
MathSciNet (http://www.ams.org/mathscinet-getitem?mr=2085)

Gao Q, Ding X, Pan F, Li W (2014) An improved sampling method of complex network. Int J Mod Phys C 25:1440007
CrossRef (http://dx.doi.org/10.1142/S0129183114400075)

García S, Molina D, Lozano M, Herrera F (2009) A study on the use of non-parametric tests for analyzing the evolutionary algorithms' behaviour: a case study on the CEC'2005 special session on real parameter optimization. J Heuristics 15:617–644
CrossRef (http://dx.doi.org/10.1007/s10732-008-9080-4)
MATH (http://www.emis.de/MATH-item?1191.68828)

Gile KJ, Handcock MS (2010) Respondent-driven sampling: an assessment of current methodology. Sociol Methodol 40:285–327
CrossRef (http://dx.doi.org/10.1111/j.1467-9531.2010.01223.x)

Gjoka M, Kurant M, Butts CT, Markopoulou A (2010) Walking in Facebook: A case study of unbiased sampling of OSNs. Proceedings IEEE INFOCOM 2010. San Diego, CA, pp 1–9
CrossRef (http://dx.doi.org/10.1109/INFCOM.2010.5462078)

Gjoka M, Butts CT, Kurant M, Markopoulou A (2011) Multigraph sampling of online social networks. IEEE J Sel Areas Commun 29:1893–1905
CrossRef (http://dx.doi.org/10.1109/JSAC.2011.111012)

Guns R, Rousseau R (2014) Recommending research collaborations using link prediction and random forest classifiers. Scientometrics 101:1461–1473
CrossRef (http://dx.doi.org/10.1007/s11192-013-1228-9)

Hall BH, Jaffe AB, Trajtenberg M (2001) The NBER patent citation data file: Lessons, insights and methodological tools. National Bureau of Economic Research

Jalali ZS, Rezvanian A, Meybodi MR (2015) A two-phase sampling algorithm for social networks. In: 2015 2nd International Conference on Knowledge-Based Engineering and Innovation (KBEI). IEEE, pp 1165–1169

Jalali ZS, Rezvanian A, Meybodi MR (2016) Social network sampling using spanning trees. Int J Mod Phys C 27:1650052
CrossRef (http://dx.doi.org/10.1142/S0129183116500522)
MathSciNet (http://www.ams.org/mathscinet-getitem?mr=3500533)

Jana R, Bagchi SB (2015) Distributional aspects of some statistics in weighted social networks. J Math Sociol 39:1–28
CrossRef (http://dx.doi.org/10.1080/0022250X.2013.866671)
MATH (http://www.emis.de/MATH-item?1320.91121)
MathSciNet (http://www.ams.org/mathscinet-getitem?mr=3301331)

Jarukasemratana S, Murata T (2015) Edge weight method for community detection on mixed scale-free networks. Int J Artif Intell Tools 24:1–24
CrossRef (http://dx.doi.org/10.1142/S0218213015400072)

Jin L, Chen Y, Hui P, et al (2011) Albatross sampling: robust and effective hybrid vertex sampling for social graphs. In: Proceedings of the 3rd ACM international workshop on MobiArch. pp 11–16

Khomami MMD, Rezvanian A, Meybodi MR (2016) Distributed learning automata-based algorithm for community detection in complex networks. Int J Mod Phys B 30:1650042
CrossRef (http://dx.doi.org/10.1142/S0217979216500429)
MathSciNet (http://www.ams.org/mathscinet-getitem?mr=3481218)

Kurant M, Markopoulou A, Thiran P (2010) On the bias of BFS (Breadth First Search). In: 2010 22nd International Teletraffic Congress (ITC). pp 1–8

Kurant M, Markopoulou A, Thiran P (2011) Towards unbiased BFS sampling. IEEE J Sel Areas Commun 29:1799–1809
CrossRef (http://dx.doi.org/10.1109/JSAC.2011.111005)

Kurant M, Gjoka M, Wang Y, et al (2012) Coarse-grained topology estimation via graph sampling. In: Proceedings of the 2012 ACM workshop on Workshop on online social networks. ACM, pp 25–30

Leskovec J, Faloutsos C (2006) Sampling from large graphs. Proceedings of the 12th ACM SIGKDD international conference on Knowledge discovery and data mining. ACM, Philadelphia, pp 631–636
CrossRef (http://dx.doi.org/10.1145/1150402.1150479)

Leskovec J, Kleinberg J, Faloutsos C (2007) Graph evolution: densification and shrinking diameters. ACM Transactions on Knowledge Discovery from Data (TKDD) 1:1–41
CrossRef (http://dx.doi.org/10.1145/1217299.1217301)

Li W, Cai X (2004) Statistical analysis of airport network of China. Phys Rev E 69:46106
CrossRef (http://dx.doi.org/10.1103/PhysRevE.69.046106)

Li M, Fan Y, Wu J, Di Z (2013a) Phase transitions in Ising model induced by weight redistribution on weighted regular networks. Int J Mod Phys B 27:1350146
CrossRef (http://dx.doi.org/10.1142/S0217979213501464)
MathSciNet (http://www.ams.org/mathscinet-getitem?mr=3109269)

Li P, Zhao Q, Wang H (2013b) A weighted local-world evolving network model based on the edge weights preferential selection. Int J Mod Phys B 27:1350039
CrossRef (http://dx.doi.org/10.1142/S0217979213500392)
MathSciNet (http://www.ams.org/mathscinet-getitem?mr=3055401)

Li Q, Zhou T, Lü L, Chen D (2014) Identifying influential spreaders by weighted LeaderRank. Phys A 404:47–55
CrossRef (http://dx.doi.org/10.1016/j.physa.2014.02.041)
MathSciNet (http://www.ams.org/mathscinet-getitem?mr=3188758)

Liben-Nowell D, Kleinberg J (2007) The link-prediction problem for social networks. J Am Soc Inform Sci Technol 58:1019–1031
CrossRef (http://dx.doi.org/10.1002/asi.20591)

Lu J, Li D (2012) Sampling online social networks by random walk. Proceedings of the First ACM International Workshop on Hot Topics on Interdisciplinary Social Networks Research. ACM, Beijing, pp 33–40
CrossRef (http://dx.doi.org/10.1145/2392622.2392628)

Lü L, Zhou T (2010) Link prediction in weighted networks: the role of weak ties. EPL (Europhysics Letters) 89:18001
CrossRef (http://dx.doi.org/10.1209/0295-5075/89/18001)

Lu Z, Sun X, Wen Y et al (2015) Algorithms and applications for community detection in weighted networks. IEEE Trans Parallel Distrib Syst 26:2916–2926
CrossRef (http://dx.doi.org/10.1109/TPDS.2014.2370031)

Luo P, Li Y, Wu C, Zhang G (2015) Toward cost-efficient sampling methods. Int J Mod Phys C 26:1550050
CrossRef (http://dx.doi.org/10.1142/S0129183115500503)

Maiya AS, Berger-Wolf TY (2010) Sampling community structure. In: Proceedings of the 19th international conference on World wide web. pp 701–710

Murai F, Ribeiro B, Towsley D, Wang P (2013) On set size distribution estimation and the characterization of large networks via sampling. IEEE J Sel Areas Commun 31:1017–1025
CrossRef (http://dx.doi.org/10.1109/JSAC.2013.130604)

Nemenyi P (1962) Distribution-free multiple comparisons. In: Biometrics. International Biometric Soc 1441 I St, Nw, Suite 700, Washington, Dc 20005-2210, p 263

Newman ME (2001) The structure of scientific collaboration networks. Proc Natl Acad Sci 98:404–409
CrossRef (http://dx.doi.org/10.1073/pnas.98.2.404)

MATH (http://www.emis.de/MATH-item?1065.00518)
MathSciNet (http://www.ams.org/mathscinet-getitem?mr=1812610)

Newman MEJ (2004) Analysis of weighted networks. Phys Rev E 70:56131
CrossRef (http://dx.doi.org/10.1103/PhysRevE.70.056131)

Newman MEJ (2006) Finding community structure in networks using the eigenvectors of matrices. Phys Rev E 74:36104
CrossRef (http://dx.doi.org/10.1103/PhysRevE.74.036104)
MathSciNet (http://www.ams.org/mathscinet-getitem?mr=2282139)

Opsahl T, Panzarasa P (2009) Clustering in weighted networks. Social networks 31:155–163
CrossRef (http://dx.doi.org/10.1016/j.socnet.2009.02.002)

Opsahl T, Agneessens F, Skvoretz J (2010) Node centrality in weighted networks: generalizing degree and shortest paths. Social Networks 32:245–251
CrossRef (http://dx.doi.org/10.1016/j.socnet.2010.03.006)

Pálovics R, Benczúr AA (2015) Temporal influence over the Last.fm social network—Springer. Social Network Analysis and Mining 5:1–12
CrossRef (http://dx.doi.org/10.1007/s13278-014-0244-y)

Papagelis M, Das G, Koudas N (2013) Sampling online social networks. IEEE Trans Knowl Data Eng 25:662–676
CrossRef (http://dx.doi.org/10.1109/TKDE.2011.254)

Park H, Moon S (2013) Sampling bias in user attribute estimation of OSNs. In: Proceedings of the 22nd international conference on World Wide Web companion. International World Wide Web Conferences Steering Committee, pp 183–184

Piña-García CA, Gu D (2013) Spiraling Facebook: an alternative Metropolis-Hastings random walk using a spiral proposal distribution. Soc Netw Anal Min 3:1403–1415
CrossRef (http://dx.doi.org/10.1007/s13278-013-0126-8)

Rejaie R, Torkjazi M, Valafar M, Willinger W (2010) Sizing up online social networks. IEEE Netw 24:32–37
CrossRef (http://dx.doi.org/10.1109/MNET.2010.5578916)

Rezvanian A, Meybodi MR (2015a) Finding maximum clique in stochastic graphs using distributed learning automata. Int J Uncertain, Fuzziness Knowl-Based Syst 23:1–31
CrossRef (http://dx.doi.org/10.1142/S0218488515500014)
MATH (http://www.emis.de/MATH-item?1323.05113)
MathSciNet (http://www.ams.org/mathscinet-getitem?mr=3312779)

Rezvanian A, Meybodi MR (2015b) Sampling social networks using shortest paths. Phys A

424:254–268
CrossRef (http://dx.doi.org/10.1016/j.physa.2015.01.030)

Rezvanian A, Meybodi MR (2016a) Stochastic graph as a model for social networks. Comput Hum Behav 64:621–640. doi:10.1016/j.chb.2016.07.032 (http://dx.doi.org/10.1016/j.chb.2016.07.032)
CrossRef (http://dx.doi.org/10.1016/j.chb.2016.07.032)

Rezvanian A, Meybodi MR (2016b) A new learning automata-based sampling algorithm for social networks. Int J Commun Syst. doi:10.1002/dac.3091:1-21 (http://dx.doi.org/10.1002/dac.3091%3A1-21)
MATH (http://www.emis.de/MATH-item?1209.68428)

Rezvanian A, Rahmati M, Meybodi MR (2014) Sampling from complex networks using distributed learning automata. Phys A 396:224–234
CrossRef (http://dx.doi.org/10.1016/j.physa.2013.11.015)

Ribeiro B, Towsley D (2010) Estimating and sampling graphs with multidimensional random walks. In: Proceedings of the 10th annual conference on Internet measurement. Melbourne, pp 390–403

Salehi M, Rabiee HR, Nabavi N, Pooya S (2011) Characterizing Twitter with Respondent-Driven Sampling. In: 2011 IEEE Ninth International Conference on Dependable, Autonomic and Secure Computing (DASC). pp 1211–1217

Salehi M, Rabiee HR, Rajabi A (2012) Sampling from complex networks with high community structures. Chaos: an Interdisciplinary. J Nonlinear Sci 22:23126
MATH (http://www.emis.de/MATH-item?1332.91096)
MathSciNet (http://www.ams.org/mathscinet-getitem?mr=3388543)

Saramaki J, Onnela J-P, Kertész J, Kaski K (2005) Characterizing motifs in weighted complex networks. Science of Complex Networks From Biology to the Internet and WWW 776:108–117

Saramäki J, Kivelä M, Onnela J-P et al (2007) Generalizations of the clustering coefficient to weighted complex networks. Phys Rev E 75:27105
CrossRef (http://dx.doi.org/10.1103/PhysRevE.75.027105)

Sett N, Singh SR, Nandi S (2016) Influence of edge weight on node proximity based link prediction methods: an empirical analysis. Neurocomputing 172:71–83
CrossRef (http://dx.doi.org/10.1016/j.neucom.2014.11.089)

Sun Y, Liu C, Zhang C-X, Zhang Z-K (2014) Epidemic spreading on weighted complex networks. Phys Lett A 378:635–640
CrossRef (http://dx.doi.org/10.1016/j.physleta.2014.01.004)
MATH (http://www.emis.de/MATH-item?1331.93190)
MathSciNet (http://www.ams.org/mathscinet-getitem?mr=3154099)

Tasgin M, Bingol HO (2012) Gossip on weighted networks. Advances in Complex Systems 15:1–18
CrossRef (http://dx.doi.org/10.1142/S0219525912500610)
MathSciNet (http://www.ams.org/mathscinet-getitem?mr=2935108)

Thi DB, Ichise R, Le B (2014) Link Prediction in Social Networks Based on Local Weighted Paths. In: Future Data and Security Engineering. Springer, pp 151–163

Tong C, Lian Y, Niu J et al (2016) A novel green algorithm for sampling complex networks. J Netw Comput Appl 59:55–62
CrossRef (http://dx.doi.org/10.1016/j.jnca.2015.05.021)

Wang S-L, Tsai Y-C, Kao H-Y et al (2013) Shortest paths anonymization on weighted graphs. Int J Software Eng Knowl Eng 23:65–79
CrossRef (http://dx.doi.org/10.1142/S0218194013400056)

Wang P, Zhao J, Lui J et al (2015) Unbiased characterization of node pairs over large graphs. ACM Transactions on Knowledge Discovery from Data (TKDD) 9:22

Yan X, Zhai L, Fan W (2013) C-index: a weighted network node centrality measure for collaboration competence. J Informetr 7:223–239
CrossRef (http://dx.doi.org/10.1016/j.joi.2012.11.004)

Yang C-L, Kung P-H, Chen C-A, Lin S-D (2013) Semantically sampling in heterogeneous social networks. In: Proceedings of the 22nd international conference on World Wide Web companion. pp 181–182

Yarlagadda R, Pinnaka S, Etinkaya EKÇ (2015) A time-evolving weighted-graph analysis of global petroleum exchange. In: 2015 7th International Workshop on Reliable Networks Design and Modeling (RNDM). IEEE, pp 266–273

Yoon S, Lee S, Yook SH, Kim Y (2007) Statistical properties of sampled networks by random walks. Phys Rev E 75:46114
CrossRef (http://dx.doi.org/10.1103/PhysRevE.75.046114)

Yoon S-H, Kim K-N, Hong J et al (2015) A community-based sampling method using DPL for online social networks. Inf Sci 306:53–69
CrossRef (http://dx.doi.org/10.1016/j.ins.2015.02.014)

Zhao SX, Rousseau R, Fred YY (2011) h-Degree as a basic measure in weighted networks. J Informetr 5:668–677
CrossRef (http://dx.doi.org/10.1016/j.joi.2011.06.005)

Zheng Y, Liu F, Gong Y-W (2014) Robustness in weighted networks with cluster structure. Mathemat Probl Eng 2014:1–8
CrossRef (http://dx.doi.org/10.1155/2014/898736)

Zhu M, Cao T, Jiang X (2014) Using clustering coefficient to construct weighted networks for supervised link prediction. Social Network Analysis and Mining 4:1–8

(2016a) The University of Florida Sparse Matrix Collection. In: The University of Florida Sparse Matrix Collection. http://www.cise.ufl.edu/research/sparse/matrices (http://www.cise.ufl.edu/research/sparse/matrices)

(2016b) Pajek datasets. In: Pajek datasets. http://vlado.fmf.uni-lj.si/pub/networks/data (http://vlado.fmf.uni-lj.si/pub/networks/data)

Copyright information

About this article

Print ISSN	Online ISSN	Publisher Name
1869-5450	1869-5469	Springer Vienna

About this journal

Reprints and Permissions