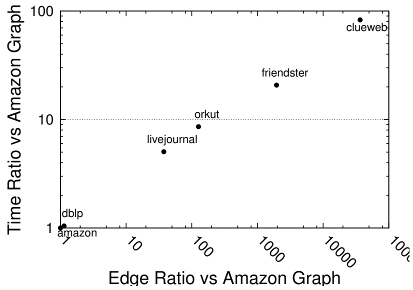

Figure 6: Running time vs size of the graph

time is only 80 times longer that of amazon. This shows that our method can scale to very large graphs with billions of nodes and edges.

7 CONCLUSIONS AND FUTURE WORK

We presented a scalable and flexible framework for finding overlapping clusters. Our empirical and theoretical findings shows that, using the local clustering structure of the ego-nets as guidance, it is possible to disentangle the complex and highly overlapping community structures of real-networks by splitting nodes into their "personas".

For future work, we would like to establish theoretical guarantees in more nuanced models by using more sophisticated partitioning algorithms. Another important direction is to adapt our methods to incremental models of computation, since real world networks are dynamic. Recent work [41] has observed a rich structure in the overlap of communities which could be analyzed to further improve our method. Finally, we believe further analysis of the structural properties of the persona graphs could yield other insights on the social network besides its clustering such as, for instance, the roles of actors in a social network [16].

ACKNOWLEDGMENTS

We thank Sergei Vassilvitskii and Bryan Perozzi for their comments and Bryan for his help with graph visualization. We thank Michele Coscia, Giulio Rossetti, Dino Pedreschi and Fosca Giannotti for making their code available and for their helpful comments.

REFERENCES

[1] B. D. Abrahao, S. Soundarajan, J. E. Hopcroft, and R. D. Kleinberg. A separability framework for analyzing community structure. *TKDD*, 2014.
[2] Y.-Y. Ahn, J. P. Bagrow, and S. Lehmann. Link communities reveal multiscale complexity in networks. *Nature*, 466(7307):761–764, 2010.
[3] L. Akoglu, M. McGlohon, and C. Faloutsos. oddball: Spotting anomalies in weighted graphs. In *PAKDD 2010*, 2010.
[4] M. Amoretti, A. Ferrari, P. Fornacciari, M. Mordonini, F. Rosi, and M. Tomaiuolo. Local-first algorithms for community detection. In *KDWeb 2016*, 2016.
[5] S. Arora, R. Ge, S. Sachdeva, and G. Schoenebeck. Finding overlapping communities in social networks: toward a rigorous approach. In *EC*, 2012.
[6] M. Balcan, C. Borgs, M. Braverman, J. T. Chayes, and S. Teng. Finding endogenously formed communities. In *SODA 2013*.
[7] S. Bandyopadhyay, G. Chowdhary, and D. Sengupta. FOCS: fast overlapped community search. *IEEE Trans. Knowl. Data Eng.*, 27(11):2974–2985, 2015.
[8] M. Bastian, S. Heymann, and M. Jacomy. Gephi: An open source software for exploring and manipulating networks, 2009.
[9] P. Boldi, M. Rosa, M. Santini, and S. Vigna. Layered label propagation: a multiresolution coordinate-free ordering for compressing social networks. In *WWW*, 2011.
[10] P. Boldi and S. Vigna. The WebGraph framework I: Compression techniques. In *WWW 2004*, pages 595–601, Manhattan, USA, 2004. ACM Press.
[11] R. Burt. *Structural Holes: The Social Structure of Competition.* Harvard Press, 1995.
[12] N. Buzun, A. Korshunov, V. Avanesov, I. Filonenko, I. Kozlov, D. Turdakov, and H. Kim. Egolp: Fast and distributed community detection in billion-node social networks. In *2014 IEEE ICDM Workshops*, pages 533–540, 2014.
[13] M. Coscia, G. Rossetti, F. Giannotti, and D. Pedreschi. Uncovering hierarchical and overlapping communities with a local-first approach. *TKDD*, 2014.
[14] A. Delis, A. Ntoulas, and P. Liakos. Scalable link community detection: A local dispersion-aware approach. In *IEEE BigData 2016*, pages 716–725, 2016.
[15] R. I. M. Dunbar and S. G. B. Roberts. Communication in social networks: Effects of kinship, network size and emotional closeness. *Personal Relationships*, 2010.
[16] D. Easley and J. Kleinberg. *Networks, crowds, and markets: Reasoning about a highly connected world.* Cambridge University Press, 2010.
[17] A. Epasto, S. Lattanzi, V. Mirrokni, I. O. Sebe, A. Taei, and S. Verma. Ego-net community mining applied to friend suggestion. *VLDB*, 9(4):324–335, 2015.
[18] M. Everett and S. P. Borgatti. Ego network betweenness. *Social Networks*, 2005.
[19] S. Fortunato. Community detection in graphs. *Physics reports*, 2010.
[20] L. C. T. Freeman. Centered graphs and the structure of ego networks. *Mathematical Social Sciences*, 1982.
[21] M. Girvan and E. J. Newman. Community structure in social and biological networks. *PNAS*, 2002.
[22] L. H. Hartwell, J. J. Hopfield, S. Leibler, and A. W. Murray. From molecular to modular cell biology. *Nature*, 402:C47–C52, 1999.
[23] K. He, Y. Sun, D. Bindel, J. E. Hopcroft, and Y. Li. Detecting overlapping communities from local spectral subspaces. In *IEEE ICDM*, pages 769–774, 2015.
[24] R. Khandekar, G. Kortsarz, and V. S. Mirrokni. On the advantage of overlapping clusters for minimizing conductance. *Algorithmica*, 69(4):844–863, 2014.
[25] R. Kumar, P. Raghavan, S. Rajagopalan, D. Sivakumar, A. Tomkins, and E. Upfal. Random graph models for the web graph. In *FOCS*, 2000.
[26] A. Lancichinetti and S. Fortunato. Benchmarks for testing community detection algorithms on directed and weighted graphs with overlapping communities. *Physical Review E*, 80(1):016118, 2009.
[27] S. Lattanzi and D. Sivakumar. Affiliation networks. In *STOC 2009*.
[28] J. Leskovec, K. J. Lang, A. Dasgupta, and M. W. Mahoney. Community structure in large networks: Natural cluster sizes and the absence of large well-defined clusters. *Internet Mathematics*, 2009.
[29] J. Leskovec, K. J. Lang, and M. W. Mahoney. Empirical comparison of algorithms for network community detection. In *WWW 2010*.
[30] R. Li, C. Wang, and K. C. Chang. User profiling in an ego network: co-profiling attributes and relationships. In *WWW*, 2014.
[31] D. Liben-Nowell and J. Kleinberg. The link-prediction problem for social networks. *journal of the Association for Information Science and Technology*, 58(7):1019–1031, 2007.
[32] W. Liu, X. Jiang, M. Pellegrini, and X. Wang. Discovering communities in complex networks by edge label propagation. *Scientific reports*, 6, 2016.
[33] J. J. McAuley and J. Leskovec. Learning to discover social circles in ego networks. In *NIPS*, 2012.
[34] A. F. McDaid, D. Greene, and N. Hurley. Normalized mutual information to evaluate overlapping community finding algorithms. Oct. 2011.
[35] G. Palla, I. Derényi, I. Farkas, and T. Vicsek. Uncovering the overlapping community structure of complex networks in nature and society. *Nature*, 2005.
[36] B. S. Rees and K. B. Gallagher. Overlapping community detection by collective friendship group inference. In *ASONAM*, 2010.
[37] P. Ronhovde and Z. Nussinov. Local resolution-limit-free potts model for community detection. *Phys. Rev. E*, 2010.
[38] S. Wasserman and K. Faust. *Social network analysis: methods and applications.* Cambridge University Press, 1994.
[39] J. Weng and B.-S. Lee. Event detection in twitter. *ICWSM*, 11:401–408, 2011.
[40] J. J. Whang, D. F. Gleich, and I. S. Dhillon. Overlapping community detection using seed set expansion. In *ACM CIKM'13*, pages 2099–2108, 2013.
[41] J. Yang and J. Leskovec. Overlapping community detection at scale: a nonnegative matrix factorization approach. In *Proceedings of the sixth ACM international conference on Web search and data mining*, pages 587–596. ACM, 2013.
[42] J. Yang and J. Leskovec. Defining and evaluating network communities based on ground-truth. *Knowledge and Information Systems*, 42(1):181–213, 2015.
[43] J. Yang, J. J. McAuley, and J. Leskovec. Detecting cohesive and 2-mode communities indirected and undirected networks. In *WSDM*, 2014.