Subgroup Discovery and Community Detection on Attributed Graphs

Martin Atzmueller

University of Kassel, Research Center for Information System Design,
Ubiquitous Data Mining Team, Chair for Knowledge and Data Engineering,
Wilhelmshöher Allee 73, 34121 Kassel, Germany

Abstract. Subgroup discovery and community detection are two approaches having been studied in different research areas like data mining and social network analysis. In this context, these techniques are especially helpful in order to provide for analytical and explorative data mining approaches, and for extracting knowledge for humans. We present an organized picture of recent research in subgroup discovery and community detection specifically focusing on attributed graphs. That is, we include complex relational graphs that are annotated with additional information, e.g., attribute information on the nodes and/or edges of the graph. Then, descriptive patterns can be extracted using a variety of techniques, ranging from structural approaches to description-based methods. This includes detecting cohesive subgroups, correlated patterns, subspace clustering, and exceptional model mining methods. We also present applications where such techniques have been successfully applied.

References

[Adnan et al. 2009] M. Adnan, R. Alhajj, J. Rokne (2009) Identifying Social Communities by Frequent Pattern Mining. Proc. 13th Intl. Conf. Inform ation Visualisation, IEEE Computer Society, Washington, DC, USA, pp. 413–418.

[Akoglu et al. 2012] L. Akoglu, H. Tong, B. Meeder, and C. Faloutsos (2012) Pics: Parameter-free Identification of Cohesive Subgroups in Large Attributed Graphs. Proc. SDM, SIAM, pp. 439–450. Omnipress

[Atzmueller 2015] Atzmueller, M (2015) Subgroup Discovery – Advanced Review. WIREs: Data Mining and Knowledge Discovery, 5(1):35–49

[Atzmueller 2007] M. Atzmueller (2007) Knowledge-Intensive Subgroup Mining – Techniques for Automatic and Interactive Discovery, Vol. 307 of Dissertations in Artificial Intelligence-Infix (Diski), IOS Press

[Atzmueller et al. 2004] M. Atzmueller, F. Puppe, H.-P. Buscher (2004) Towards Knowledge-Intensive Subgroup Discovery, Proc. LWA 2004, pp. 117–123.

[Atzmueller & Puppe 2006] M. Atzmueller and F. Puppe (2006) SD-Map - A Fast Algorithm for Exhaustive Subgroup Discovery. Proc. 10th European Conf. on Principles and Practice of Knowledge Discovery in Databases (PKDD 2006), pp. 6-17, Heidelberg, Germany. Springer Verlag

[Atzmueller et al. 2005] M. Atzmueller, J. Baumeister, A. Hemsing, E.-J. Richter, and F. Puppe (2005) Subgroup Mining for Interactive Knowledge Refinement. In Proc. 10th Conference on Artificial Intelligence in Medicine AIME 05), LNAI 3581, pp. 453-462, Heidelberg, Germany, Springer Verlag.

[Atzmueller et al. 2005] M. Atzmueller, F. Puppe, and H.-P. Buscher (2005) Pro-filing Examiners using Intelligent Subgroup Mining. In Proc. 10th International Workshop on Intelligent Data Analysis in Medicine and Pharmacology (IDAMAP-2005), pp. 46-51, Aberdeen, Scotland

[Atzmueller & Puppe 2008] M. Atzmueller and F. Puppe (2008) A Case-Based Approach for Characterization and Analysis of Subgroup Patterns. Journal of Applied Intelligence, 28(3):210-221

[Atzmueller & Hilgenberg 2013] M. Atzmueller and K. Hilgenberg (2013) Towards Capturing Social Interactions with SDCF: An Extensible Framework for Mobile Sensing and Ubiquitous Data Collection. In Proc. 4th International Workshop on Modeling Social Media (MSM 2013), Hypertext 2013, New York, NY, US. ACM Press.

[Atzmueller & Lemmerich 2012] M. Atzmueller and F. Lemmerich (2012) VIKAMINE - Open-Source Subgroup Discovery, Pattern Mining, and Analytics. In Proc. ECML/PKDD 2012: European Conference on Machine Learning and Principles and Practice of Knowledge Discovery in Databases, Heidelberg, Germany. Springer Verlag.

[Atzmueller & Puppe 2005] M. Atzmueller and F. Puppe (2005) Semi-Automatic Visual Subgroup Mining using VIKAMINE. Journal of Universal Computer Science, 11(11):1752-1765, 2005.

[Atzmueller & Lemmerich 2009] M. Atzmueller, F. Lemmerich (2009) Fast Subgroup Discovery for Continuous Target Concepts. Proc. International Symposium on Methodologies for Intelligent Systems, Vol. 5722 of LNCS, Springer, Berlin, pp. 1–15.

[Atzmueller et al. 2012] M. Atzmueller, S. Doerfel, A. Hotho, F. Mitzlaff, and G. Stumme (2012) Face-to-Face Contacts at a Conference: Dynamics of Communities and Roles. In Modeling and Mining Ubiquitous Social Media, volume 7472 of LNAI. Springer Verlag, Heidelberg, Germany

[Atzmueller & Lemmerich 2013] M. Atzmueller and F. Lemmerich (2013) Exploratory Pattern Mining on Social Media using Geo-References and Social Tagging Information. IJWS, 2(1/2)

[Atzmueller & Mitzlaff 2011] M. Atzmueller and F. Mitzlaff (2011) Efficient Descriptive Community Mining. Proc. 24th International FLAIRS Conference, pages 459-464, Palo Alto, CA, USA. AAAI Press.

[Atzmueller et al. 2015] M. Atzmueller, S. Doerfel, and F. Mitzlaff (2015) Description-Oriented Community Detection using Exhaustive Subgroup Discovery. Information Sciences. http://dx.doi.org/10.1016/j.ins.2015.05.008.

[Atzmueller et al. 2009] M. Atzmueller, F. Lemmerich, B. Krause, and A. Hotho (2009) Who are the Spammers? Understandable Local Patterns for Concept Description. In Proc. 7th Conference on Computer Methods and Systems, Krakow, Poland. Oprogramowanie Nauko-Techniczne.

[Berlingerio et al. 2013] M. Berlingerio, F. Pinelli, and F. Calabrese (2013) ABA-CUS: Apriori-BAsed Community discovery in mUltidimensional networkS. Data Mining and Knowledge Discovery, Springer, 27(3).

[Boden et al. 2012] B. Boden, S. Günnemann, H. Hoffmann, and T. Seidl (2012) Mining Coherent Subgraphs in Multi-Layer Graphs with Edge Labels. Proc. 18th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. New York, USA: ACM Press

[Bothorel et al. 2015] C. Bothorel, J. D. Cruz, M. Magnani, B. Micenkova (2015) Clustering Attributed Graphs: Models, Measures and Methods. arXiv:1501.01676

[Bringmann et al. 2011] B. Bringmann, S. Nijssen, and A. Zimmermann (2011) Pattern-based Classification: A Unifying Perspective. arXiv:1111.6191

[Clauset et al. 2004] A. Clauset, M. E. J. Newman, C. Moore (2004) Finding Community Structure in Very Large Networks. arXiv:cond-mat/0408187

[Cruz et al. 2011] J. D. Cruz, C. Bothorel, F. and Poulet (2011) Entropy Based Community Detection in Augmented Social Networks. Computational Aspects of Social Networks, pp. 163-168

[Dang & Viennet 2012] T. A. Dang and E. Viennet (2012) Community Detection Based on Structural and Attribute Similarities. Proc. International Conference on Digital Society (ICDS), pp. 7-14

[Duivestein et al. 2015] W. Duivesteijn, A.J. Feelders, and A. Knobbe (2015) Exceptional Model Mining - Supervised Descriptive Local Pattern Mining with Complex Target Concepts. Data Mining and Knowledge Discovery

[Fortunato 2010] S. Fortunato (2010) Community Detection in Graphs, Physics Reports 486 (3-5)

[Freeman 1978] L. Freeman (1978) Segregation In Social Networks, Sociological Methods & Research 6 (4)

[Ge et al. 2008] R. Ge, M. Ester, B. J. Gao, Z. Hu, B. Bhattacharya, and B. Ben-Moshe (2008) Joint Cluster Analysis of Attribute Data and Relationship Data: The Connected k-Center Problem, Algorithms and Applications. Acm Trans. Knowl. Discov. Data, 2(2)

[Girvan & Newman 2002] M. Girvan, M. E. J. Newman (2002) Community Structure in Social and Biological Networks, PNAS 99 (12)

[Günnemann et al. 2013] S. Günnemann, I. Färber, B. Boden, T. Seidl (2013) GAMer: A Synthesis of Subspace Clustering and Dense Subgraph Mining. Knowledge and Information Systems (KAIS), Springer

[Kannan et al. 2004] R. Kannan, S. Vempala, A. Vetta (2004) On Clustering: Good, Bad and Spectral. Journal of the ACM, 51(3)

[Kloesgen 1996] Kloesgen, W. (1996) Explora: A Multipattern and Multistrategy Discovery Assistant. In Fayyad, U. M., Piatetsky-Shapiro, G., Smyth, P., and Uthurusamy, R., editors, Advances in Knowledge Discovery and Data Mining, pp. 249–271. AAAI Press.

[Lancichinetti 2009] A. Lancichinetti, S. Fortunato (2009) Community Detection Algorithms: A Comparative Analysis. arXiv:0908.1062

[Lazarsfield & Merton 1954] P. F. Lazarsfeld, R. K. Merton (1954) Friendship as a Social Process: A Substantive and Methodological Analysis. Freedom and Control in Modern Society, 18(1), 18-66

[Leman et al. 2008] D. Leman, A. Feelders, and A. Knobbe (2008). Exceptional Model Mining. In Proc. European Conference on Machine Learning and Principles and Practice of Knowledge Discovery in Databases, volume 5212 of Lecture Notes in Computer Science, pages 1–16. Springer.

[Lemmerich et al. 2012] F. Lemmerich, M. Becker, and M. Atzmueller (2012) Generic Pattern Trees for Exhaustive Exceptional Model Mining. In Proc. ECML/PKDD, Heidelberg, Germany. Springer

[Leskovec et al. 2010] J. Leskovec, K. J. Lang, and M. Mahoney (2010) Empirical Comparison of Algorithms for Network Community Detection. Proc. 19th International Conference on World Wide Web, pp. 631-640. ACM

[McPherson et al. 2011] M. McPherson, L. Smith-Lovin, and J. M. Cook (2001) Birds of a Feather: Homophily in Social Networks. Annual Review of Sociology, 415-444

[Mitzlaff et al. 2011] F. Mitzlaff, M. Atzmueller, D. Benz, A. Hotho, and G. Stumme (2011) Community Assessment using Evidence Networks. In Analysis of Social Media and Ubiquitous Data, volume 6904 of LNAI

[Mitzlaff et al. 2013] F. Mitzlaff, M. Atzmueller, D. Benz, A. Hotho, and G. Stumme (2013) User-Relatedness and Community Structure in Social Interaction Networks. CoRR/abs, 1309.3888

[Moser et al. 2009] F. Moser, R. Colak, A. Rafiey, and M. Ester (2009) Mining Cohesive Patterns from Graphs with Feature Vectors. Proc. SDM (Vol. 9), pp. 593-604.

[Newman 2004] M. E. Newman (2004). Detecting community structure in networks. The European Physical Journal B-Condensed Matter and Complex Systems, 38(2), 321-330.

[Newman 2006] M. E. Newman 2006) Modularity and Community Structure in Networks. PNAS, 103(23), 8577-8582.

[Palla et al. 2005] G. Palla, I. Derenyi, I. Farkas, and T. Vicsek (2005) Uncovering the Overlapping Community Structure of Complex Networks in Nature and Society. Nature, 435(7043), 814-818

[Pool et al. 2014] S. Pool, F. Bonchi, M. van Leeuwen (2014) Description-driven Community Detection, Transactions on Intelligent Systems and Technology 5 (2)

[Psorakis et al. 2011] I. Psorakis, S. Roberts, M. Ebden, and B. Sheldon. Overlapping Community Detection using Bayesian Non-Negative Matrix Factorization. Phys. Rev. E 83, 066114

[Puppe et al. 2008] F. Puppe, M. Atzmueller, G. Buscher, M. Huettig, H. Lührs, and H.-P. Buscher (2008) Application and Evaluation of a Medical Knowledge-System in Sonography (SonoConsult). In Proc. 18th European Conference on Artificial Intelligence (ECAI 20008), pp. 683-687

[Ruan et al. 2013] Y. Ruan, D. Fuhry, and S. Parthasarathy (2013). Efficient Community Detection in Large Networks Using Content and Links. Proc. 22nd International Conference on World Wide Web, pp. 1089–1098, ACM.

[Tang & Liu 2010] L. Tang and H. Liu (2010) Community Detection and Mining in Social Media. Synthesis Lectures on Data Mining and Knowledge Discovery, 2(1), 1-137. Morgan & Claypool Publishers

[Steinhaeuser & Chawla 2008] K. Steinhaeuser, N. V. Chawla (2008) Community Detection in a Large Real-World Social Network. Social Computing, Behavioral Modeling, and Prediction, pp. 168–175, Springer

[Silva et al. 2012] A. Silva, W. Meira Jr., and M. J. Zaki (2010) Structural Correlation Pattern Mining for Large Graphs. Proc. Workshop on Mining and Learning with Graphs. MLG '10, pp. 119–126. New York, NY, USA: ACM.

[Smith et al. 2014] L. M. Smith, L. Zhu, K. Lerman, and A. G. Percus. Partitioning Networks with Node Attributes by Compressing Information Flow. arXiv:1405.4332

[Scholz et al. 2013] C. Scholz, M. Atzmueller, A. Barrat, C. Cattuto, and G. Stumme (2013). New Insights and Methods For Predicting Face-To-Face Contacts. Proc. 7th Intl. AAAI Conference on Weblogs and Social Media, Palo Alto, CA, USA, AAAI Press.

[Wassermann & Faust 1994] S. Wasserman, and K. Faust (1994) Social Network Analysis: Methods and Applications. Structural Analysis in the Social Sciences. Cambridge University Press, 1 edition.

[Wrobel 1997] S. Wrobel (1997) An Algorithm for Multi-Relational Discovery of Subgroups. In Proc. 1st Europ. Symp. Principles of Data Mining and Knowledge Discovery, pages 78–87, Heidelberg, Germany. Springer Verlag.

[Xie et al. 2013] J. Xie, S. Kelley, and B. K. Szymanski (2013) Overlapping Community Detection in Networks: The State-of-the-art and Comparative Study. ACM Comput. Surv., 45(4):43:1–43:35.

[Xu et al. 2012] Z. Xu, Y. Ke, Y. Wang, H. Cheng, and J. Cheng (2012) A Model-based Approach to Attributed Graph Clustering. Proc. ACM International Conference on Management of Data. SIGMOD '12, pp. 505–516, New York, NY, USA. ACM.

[Yang et al. 2013] J. Yang, J. McAuley, and J. Leskovec (2013) Community Detection in Networks with Node Attributes. Proc. IEEE International Conference on Data Mining (ICDM), pp. 1151–1156. IEEE Press, Washington, DC, USA

[Zachary, 1977] W. W. Zachary (1977) An Information Flow Model for Conflict and Fission in Small Groups. Journal of Anthropological Research, 452-473.

[Zhou et al. 2009] Y. Zhou, H. Cheng, and J. X. Yu (2009) Graph Clustering Based on Structural/Attribute Similarities. Proc. VLDB Endow., 2(1), 718–729.