Mining Attributed Networks

Rushed Kanawati and 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

Tutorial Abstract. Research in modeling, analyzing and mining large-scale networks has attracted an increasing effort in the last few years. A major trend of work in network modeling and mining concerns analyzing homogeneous static networks (i. e., one snapshot of a network). However, in real world settings, networks are often dynamic, heterogeneous, and both nodes and links can be described by a set of attributes. To cope with that problem, two current main trends are explored in our tutorial: multiplex networks and attributed networks analysis approaches.

A multiplex network is often represented as a multi-layer network composed of a set of nodes related to each other with different types of relations. This representation is much richer than simple monoplex networks. However, this poses the challenge to provide adequate answers to all basic network analysis tasks that have been studied and provided in the recent few years. Almost all work in the field of multiplex and attributed network analysis are based on transforming the problem, in a way or another to the classical case of homogeneous network analysis. In this tutorial, we focus mainly on new approaches that extend existing approaches to new rich representations of multiplex and attributed networks, and provide an organized picture on analytical methods, also including exemplary applications.

References

[Blondel et al. 2008] V. Blondel, J. Guillaume, and E. Lefebvre (2008) Fast Unfolding of Communities in Large Networks. Journal of Statistical Mechanics: Theory and Experiment

[Clauset 2005] A. Clauset (2005) Finding Local Community Structure in Networks. Physical Review E

[Chen et al. 2009] J. Chen, O. Zaïane, and R. Goebel(2009) Local community identification in social networks, ASONAM, 2009, pp. 237–242.

[Good et al. 2010] B. H. Good, Y.-A. de Montjoye, and A. Clauset (2010) The Performance of Modularity Maximization in Practical Contexts. Physical Review E, no. 81, 046106.

[Lancichinetti et al. 2011] A. Lancichinetti, and S. Fortunato (2011) Limits of Modularity Maximization in Community Detection CoRR abs/1107.1.

In: DSAA 2017, Tokyo, Japan, 2017

[Lue et al. 2008] F Luo, J. Wang, and E. Promislow (2008) Exploring Local Community Structures in Large Networks. Web Intelligence and Agent Systems 6, no. 4, 387–400.

[Pizzuti 2012] C. Pizzuti (2012) A Multiobjective Genetic Algorithm to Find Communities in Complex Networks, IEEE Trans. Evolutionary Computation 16, no. 3, 418–430.

[Pons & Latapy 2006] P. Pons, and M. Latapy (2006) Computing Communities in Large Networks using Random Walks, J. Graph Algorithms Appl. 10, no. 2, 191–218.

[Yang & Leskovec 2012] J. Yang, and J. Leskovec (2012) Defining and Evaluating Network Communities based on Ground-Truth. ICDM, IEEE Computer Society, pp. 745–754.

[Amelia & Pizzuti 2014] A. Amelio, and C. Pizzuti (2014) Community Detection in Multidimensional Networks. IEEE 26th International Conference on Tools with Artificial Intelligence, pp. 352–359.

[Berlingerio et al. 2011] M. Berlingerio, M. Coscia, and F. Giannotti (2011) Finding and Characterizing Communities in Multidimensional Networks. ASONAM, IEEE Computer Society, pp. 490–494.

[Battiston et al. 2013] F. Battiston, V. Nicosia, and V. Latora (2013) Metrics for the Analysis of Multiplex Networks. CoRR abs/1308.3182

[Dwork et al. 2001] C. Dwork, R. Kumar, M. Naor, and D. Sivakumar (2001) Rank Aggregation Methods for the Web, WWW, pp. 613–622.

[De Domenico et al. 2015] M. De Domenico, A. Lancichinetti, A. Arenas, and M. Rosvall (2015) Identifying Modular Flows on Multilayer Networks Reveals Highly Overlapping Organization in Social Systems. Phys. Rev 5

[Kanawati 2014] R. Kanawati (2014) Seed-Centric Approaches for Community Detection in Complex Networks, 6th International Conference on Social Computing and Social Media, vol. LNCS 8531, Springer, pp. 197–208.

[Kuncheva & Montana 2015] Z Kuncheva, and G. Montana (2015) Community Detection in Multiplex Networks using Locally Adaptive Random Walks, MANEM 2 workshop - Proceedings of ASONAM 2015 (Paris)

[Mucha et al. 2010] P. J. Mucha, T. Richardson, K. Macon, M. A. Porter, and J.-P. Onnela (2010) Community Structure in Time-Dependent, Multiscale, and Multiplex Networks, Science 328), no. 5980, 876–878.

[Pujari & Kanawati 2012] M. Pujari and R. Kanawati (2012) Supervised Rank Aggregation Approach for Link Prediction in Complex Networks. WWW (Companion Volume), ACM, 2012, pp. 1189–1196.

[Strehl & Gosh 2003] A. Strehl and J. Ghosh (2003) Cluster Ensembles: A Knowledge Reuse Framework for Combining Multiple Partitions. The Journal of Machine Learning Research **3** (2003), 583–617.

[Yakoubi & Kanawati 2014] Zied Yakoubi and Rushed Kanawati, *Licod: Leader-driven approaches for community detection*, Vietnam Journal of Computer Science 1, no. 4, 241–256.

[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. 2006] M. Atzmueller, J. Baumeister, and F. Puppe (2006) Introspective Subgroup Analysis for Interactive Knowledge Refinement. Proc. FLAIRS, AAAI Press, Palo Alto, CA, USA, pp. 402-407

[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 & 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 & 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. 2016] M. Atzmueller, S. Doerfel, and F. Mitzlaff (2016) Description-Oriented Community Detection using Exhaustive Subgroup Discovery. Information Sciences. 329, 965-984

[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) ABACUS: 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)

[Kibanov et al. 2014] M. Kibanov, M. Atzmueller, C. Scholz, and G. Stumme (2014) Temporal Evolution of Contacts and Communities in Networks of Face-to-Face Human Interactions. Science China, 57

[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

[Lemmerich et al. 2010] F. Lemmerich, M. Rohlfs, and M. Atzmueller (2012) Fast Discovery of Relevant Subgroup Patterns. Proc. 23rd FLAIRS Conference, AAAI Press, New York, NY, USA.

[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. 2014] F. Mitzlaff, M. Atzmueller, A. Hotho, and G. Stumme (2014) The Social Distributional Hypothesis: A Pragmatic Proxy for Homophily in Online Social Networks. Social Network Analysis and Mining 4(1), 216

[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.

[Akbari et al. 2015] E. Akbari, H. M. Dahlan, R. Ibrahim, and H. Alizadeh, (2015) *Hierarchical cluster ensemble selection*, Engineering Applications of Artificial Intelligence 39, 146–156.

[Azimi & Fern 2009] J. Azimi, and X. Fern (2009) Adaptive Cluster Ensemble Selection. IJCAI, pp. 992–997.

[Aggarwal & Chandan 2014] C. C. Aggarwal, and C. K. Reddy (eds.) (2014) Data Clustering: Algorithms and Applications. CRC Press, 2014.

[Atzmueller et al. 2017] M. Atzmueller, A. Schmidt, B. Kloepper, and D. Arnu (2017) HypGraphs: An Approach for Analysis and Assessment of Graph-Based and Sequential Hypotheses. New Frontiers in Mining Complex Patterns. Post-proceedings NFMCP 2016 (Heidelberg, Germany), LNAI, Springer.

[Atzmueller 2016] M. Atzmueller (2016) Detecting Community Patterns Capturing Exceptional Link Trails. Proc. IEEE/ACM ASONAM (Boston, MA, USA), IEEE Press.

[Atzmueller et al. 2016] M. Atzmueller, A- Schmidt, and M. Kibanov (2016) DASH-Trails: An Approach for Modeling and Analysis of Distribution-Adapted Sequential Hypotheses and Trails. Proc. WWW 2016 (Companion)

[Atzmueller et al. 2014] M. Atzmueller, M. Becker, M. Kibanov, C. Scholz, S. Doerfel, A. Hotho, B.-E. Macek, F. Mitzlaff, J. Mueller, and G. Stumme (2014) Ubicon and its Applications for Ubiquitous Social Computing. New Review of Hypermedia and Multimedia, (20)1:53–77

[Atzmueller & Roth-Berghofer 2010] M. Atzmueller, and T. Roth-Berghofer (2010) The Mining and Analysis Continuum of Explaining Uncovered, Proc. 30th SGAI International Conference on Artificial Intelligence (AI-2010)

[Boden et al. 2017] B. Boden, S. Günnemann, H. Hoffmann, and T. Seidl (2017) Mimag: Mining Coherent Subgraphs in Multi-Layer Graphs with Edge Labels. Knowledge and Information Systems 50, no. 2, 417–446.

[Manlio et al. 2014] M. De Domenico, M. A. Porter, and A. Arenas (2014) Multilayer Analysis and Visualization of Networks. J. Complex Networks 10.

[Espin-Noboa 2017] L. Espín-Noboa, F. Lemmerich, M. Strohmaier, and P. Singer (2017) Janus: A Hypothesis-Driven Bayesian Approach for Understanding Edge Formation in Attributed Multigraphs, Applied Network Science 2, no. 1, 16.

[Falih & Kanawati] I. Falih, and R. Kanawati (2015) Muna: A Multiplex Network Analysis Library. The 2015 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (Paris), pp. 757–760.

[Fern & Lin 2008] X. Z. Fern, and W. Lin (2008) Cluster Ensemble Selection. Statistical Analysis and Data Mining 1, no. 3, 128–141.

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

[Kanawati 2015] R. Kanawati (2015) Empirical Evaluation of Applying Ensemble Methods to Ego-centered Community Identification in Complex Networks. Neurocomputing 150, B, 417–427.

[Meila 2003] M. Meila (2003) Comparing Clusterings by the Variation of Onformation, COLT (Bernhard Schölkopf and Manfred K. Warmuth, eds.), Lecture Notes in Computer Science, vol. 2777, Springer, pp. 173–187.

[Murilo et al. 2013] M. C. Naldi, A. C. P. L. F. Carvalho, and R. J. G. B. Campello (2013) Cluster Ensemble Selection Based on Relative Validity Indexes, Data Min. Knowl. Discov. 27, no. 2, 259–289.

[Papadopoulos et al. 2017] A. Papadopoulos, G. Pallis, and M. D Dikaiakos (2017) Weighted Clustering of Attributed Multi-Graphs. Computing 99, no. 9, 813–840.

[Roth-Berghofer et al. 2007] T. Roth-Berghofer, S. Schulz, D. Leake, and D. Bahls (2007) Explanation-Aware Computing, AI Magazine 28, no. 4.