Algorithms for analyzing evolving networks on the Dark Web & in science

Boekhout, H.D.

7 Conclusion

This chapter concludes the thesis, titled "Algorithms for Analyzing Evolving Networks on the Dark Web & in Science", in which we have introduced various methodological innovations in order to study a Dark Web cryptomarket and the scientific research system over time. We summarize the conclusions of the presented research and answer the questions posed in Section 1.5. First, we mention a number of methodological and algorithmic advances that were made in this regard in Section 7.1. After that, we discuss several findings related to our analyses of communication networks on the dark web in Section 7.2 and scientific collaboration networks in Section 7.3.

7.1 Methodological and algorithmic advances

A substantial methodological advance presented in this thesis, is the proposed algorithm for fast *temporal maximal clique* enumeration introduced in Chapter 4. These temporal cliques define for a set of nodes a timespan wherein they are fully connected, with a given minimum weight, during each time interval of a given length. Enumerating all such temporal cliques allow us to study the temporal group dynamics in a system. Similar to an existing algorithm [8], the presented algorithm consists of a stretching and a bulking phase. However, by changing the approach to the stretching phase, our new algorithm gained the ability to deal not only with temporal networks, but also with weighted temporal networks. Additionally, the new approach guarantees a linear runtime with respect to the number of temporal edges of the input network, for this phase of the algorithm. Furthermore, a combination of a strict tree search order and efficient pruning of branches of the search tree, was shown to lead to substantial runtime improvements compared to existing algorithms for both real-world and synthetic temporal networks. Additionally, our proposed algorithm was shown to scale better for synthetic temporal networks with higher structural, i.e. static, and temporal density. Consequently, the introduced algorithm allows for the processing of larger and more complex networks than before. As the tree search order was found to directly affect the total amount of pruning, one future algorithmic advance could be found in developing methods of choosing an effective search order through custom node labeling. Another advantage comes in the form of parallelization, to which the algorithm design is particularly well suited.

To generate the set of synthetic temporal networks with varying structural and temporal densities for our experiments in Chapter 4, we introduced a new method for generating such networks. In order to control for the structural density, our method relies on the Erdös-Rényi and Barabasi-Albert random graph models to produce the underlying static network. At least one temporal interaction is presumed to exist for each edge in the static network and a Poisson process is used to simulate the number of additional temporal interactions. By varying the λ parameter we are then able simulate different temporal densities. The actual timestamps were subsequently chosen close to existing or neighboring timestamps, using normal distributions. This approach was used to enforce temporal clustering and simulate group activity and temporal burstiness. Our proposed method is well suited to generating a set of synthetic networks in which the structural and the temporal densities can be varied separately. However, we do not claim this proposed method as one of our methodological innovations, because we have not yet investigated how well the resulting networks represent real-world temporal networks. We leave this for future work.

The second methodological innovation, as presented in Chapter 6, introduced a new procedure for identifying *persistent scientific teams*. These persistent scientific teams are groups of scientific authors, along with a given timespan, where each pair of authors is in a persistent collaboration during the team's full timespan. Two authors are considered to be in a persistent collaboration if they co-authored at least three publications within a five year period. Consequently, the process of identifying persistent scientific teams consists of three steps: (1) constructing the temporal co-authorship network from publication data; (2) transforming this into a persistent collaboration network; and (3) extracting persistent teams from the persistent collaboration network through temporal maximal clique enumeration. Here, the final step can utilize the efficient algorithm introduced as our first methodological innovation. Where scientific team research is often limited to teams associated with a single publication, grant, institute or department, this new procedure allows us to identify scientific teams that capture the fluid and interdisciplinary nature of modern science. Whereas we focus on one global definition of persistent collaboration in our efforts to identify scientific teams, future work could instead explore field-specific definitions that capture the differences in the expected productivity for each field.

The third methodological innovation introduced the new spatial unit *scientific city* in Chapter 5, for use in research of the scientific research system. This unit defines a spatial distance that goes beyond the organization yet retains the idea of easier collaboration due to short travel times. At the same time, this spatial unit tries to avoid the common pitfall in traditional city definitions where huge global cities with extensive infrastructure end up being compared either against small cities and towns directly or against agglomerations of cities and towns that may have poor interconnected infrastructure. Each scientific city is centered around a scientific organization or institution that is the most productive scientific entity, in terms of scientific output, within an eight kilometer radius of itself. The scientific city is subsequently considered to be the collection of all scientific organizations and institutions within the eight kilometer radius of the center. To reduce the number of cities with very low scientific output, we merged cities with fewer than ten publications into the closest scientific city with more than a hundred publications (within 30 kilometers). A manual inspection of a sample of regions, including “de Randstad” in the Netherlands, which is a notoriously difficult region for agglomeration, found that this approach worked well throughout most regions of the world, often dividing world cities into their named districts. Unfortunately, Chinese cities such as Beijing however, do not allow for segregation into smaller scientific cities as the addresses listed on publications tend to be at the municipality level. This issue would need to be addressed in future work to make the scientific city unit truly globally applicable.

The fourth and final methodological innovation presented in this thesis, is the proposed algorithm for generating *Evolving Degree Respecting Rewired* (EDRR) networks introduced in Chapter 5. EDRR networks are meant to simulate the rewiring, i.e., changes in edge connections, that occurs in an evolving network from one timestep to the next. This approach to random rewiring of the network relies on the assumption that the network under consideration is largely static between any two subsequent timesteps. When this assumption holds, such as for the scientific city co-authorship network studied in Chapter 5, EDRR networks allow us to compare the true changes in the network evolution to a null model that simulates these changes while indirectly retaining the new degree distribution as good as possible. It attempts to accomplish this by applying the same number of edge deletions and edge additions as that occur in the real network. For these deletions and additions, edges (i.e., node pairs) are selected at random from those node pairs that have highly similar degrees as those to which the true deletions and additions were applied. A comparison with this null model can highlight where in a network (i.e., in the core, periphery, etc.) changes occur more frequently than expected. However, note that this statement is based on the EDRR model design principles and is not yet backed up by thorough testing.

7.2 Dark Web Market findings

In Chapter 2 we extracted dark web forum and market data from raw source files. During this process a multitude of data quality issues were identified and resolved. We found data incompleteness, especially in terms of missing data records, to be the most limiting factor in resolving other data quality issues. Data inconsistencies and data currency issues were often easily resolved, or understood as actual changes, when surrounding data records were complete, while duplicate data actually served to bolster otherwise incomplete records. This showed that the measures that were taken, to speed-up the scraping process of these raw sources files by preventing the acquisition of duplicate data with respect to previous scrapes, actually harms the ability to resolve data quality issues afterwards. Thus, the intentional inclusion of duplicate data between scrapes, should be seriously considered in any forum scraping effort.

On the forum side, topic and post information was relatively complete. Most missing information resided on parts of the forum that were hidden from most users, with an estimated 10% of topics and 8% of posts hidden. On the other hand, user information was less complete, with information missing on approximately 12% of users. On the market side, listing and vendor details were more incomplete. Over 40% of listings and over 25% of vendors had information missing that could be crucial to in-depth analysis of the market. Fortunately, only 6% of vendors had missing sales information, enabling us to investigate the "success" of most vendors.

In Chapter 3 we investigated the relationship between vendors (and their success) and the role they played on the forum. Note that approximately 60% of vendor usernames could be directly matched to corresponding forum usernames (Chapter 2). To capture the role users played on the forum we considered two types of measures: (1) forum activity indicators, i.e., post activity, topics started, and topic engagement; and (2) network centrality measures, i.e., degree, harmonic closeness centrality, betweenness centrality, and PageRank. To this end, we constructed monthly communication networks from the extracted forum data. To ensure a meaningful network representation, we considered a communication tie to only exist if two posts followed each other within a reasonable time frame and without too many interceding posts. This ensured that we could more confidently consider a tie to represent actual communication of information between two users. Monthly networks, including all data prior and including the given month, were constructed so that the relationship between vendor success and their forum activity could be studied over time. Note that because the relationship between posts being placed and other users reading them is asynchronous, we considered only measures that do not rely on time-respecting paths. After all, there is no requirement for preceding posts to have been read in temporal order.

Topic engagement, which measures the extent to which other users posted in topics started by the user under consideration, was shown to be the best indicator of them being a vendor as well as their success as a vendor. In this, topic engagement was closely followed by betweenness centrality, indicating an ability of (successful) vendors to connect otherwise disparate users, i.e., to often be central in discussions on the forum. Notably, betweenness centrality is a better indicator of (successful) vendors when considering vendors with relatively low forum activity. Therefore, betweenness centrality may provide added value in law enforcement effort to identify successful vendors. Furthermore, both topic engagement and betweenness centrality were shown to be a decent predictor of future vendor success, with betweenness centrality having the strongest indications. Additionally, betweenness centrality was shown to be a good indicator of other crucial positions for the proper functioning of the forum, and consequently its marketplace, such as moderators and administrators.

Future work investigating other, and more recent, cryptomarkets is needed to establish whether our findings can be generalized. Moreover, such future work may help to establish (a range of) sensible and generally applicable parameters for communication network extraction. This would be useful to law enforcement, as their application of our findings would most likely focus on settings where no sales information is known, i.e., where parameters cannot be tuned ad hoc. Finally, while we did experiment with a wider selection of measures, our analysis was not exhaustive in terms of considered network measures. Additionally, while we had good reason to focus on static measures, time-respecting measures may provide unexpected new insights. Therefore, future work may want to consider other, potentially time-respecting path based, measures in addition to betweenness centrality.

7.3 Scientific research system findings

Chapter 5 analyzed the evolution of a time-sliced scientific city collaboration network over time and how this differed from simulated evolutions using EDRR networks. Comparing the rankings of scientific cities for a variety of centrality measures from one time-slice to the next, we observed that the rankings have become more stable over time. A comparison of ranking stability with generated EDRR networks for closeness and betweenness centrality, showed that the EDRR based rankings suffered similar amounts of changes as those based on the real-world network, but that the changes in rank were more substantial in general. These findings combined, imply that changes in the scientific city collaboration network, such as new city co-authorships, occur more frequently between “close” cities in the network periphery.

Finally, in Chapter 6 we considered the relationship between scientific team freshness, persistence, and the scientific success of the team. Persistent scientific teams were shown to be prevalent in the majority of scientific output and overrepresented among high-impact research. Notably, teams benefit from institutional and geographic diversity until coordination costs start to outweigh gains. Furthermore, we observed that teams tend to produce their highly cited publications early in the collaboration, highlighting the major role freshness plays in the success of a team. Moreover, analyzing how team members engaging in other persistent collaborations affect the team's success, we observed that freshness impulses, from newly started teams, contribute to the (continued) success of the team. However, too many such freshness impulses, i.e., a lack of focus and increased coordination costs, can hinder success. Thus, in the balance between persistence and freshness, freshness appears to be the more important factor. However, we also observed that a team's persistence can greatly reduce the early coordination costs of subsequently formed teams, thereby contributing to early team success.

While we found interesting global patterns in Chapter 6 with respect to the influence of team freshness, persistence, and openness on team success, we also observed geographical differences in the prevalence of teams in the scientific output of countries. It would be highly informative for policy makers, for future studies to investigate how the observed patterns differ by country and by field. Additionally, early results showed that (international) author mobility may well play an important role in the relationship between team openness and success. Therefore, it would be interesting for future research to explore the role played by author mobility in this context.

Throughout this thesis, time and the proper use of temporal information has been a central theme. Yet, temporal information has played many different roles: from aiding in data quality resolution (Chapter 2); to use in network snapshot creation (Chapters 2, 3 and 5); to allowing the study of data both before and after a moment in time (Chapter 3); to incorporation in a network structure, e.g., cliques, and their enumeration (Chapter 4); and, finally to its incorporation in a unit of analysis, e.g., scientific teams, and their study over time (Chapter 6). Consequently, employing temporal information can be key to answering research questions and gaining new insights, as long as its use is well considered and appropriate to the situation.

Bibliography

[1] The sociopatterns data repository of physical proximity and face-to-face contacts of individuals in numerous real-world environments. `http://www.sociopatterns.org`. Accessed: 2024-06-10.

[2] L. Armona. Measuring the demand effects of formal and informal communication: Evidence from online markets for illicit drugs. *arXiv preprint arXiv:1802.08778*, 2018.

[3] P. Ayton and I. Fischer. The hot hand fallacy and the gambler's fallacy: Two faces of subjective randomness? *Memory & cognition*, 32:1369–1378, 2004.

[4] X. Bai, H. Pan, J. Hou, T. Guo, I. Lee, and F. Xia. Quantifying success in science: An overview. *IEEE Access*, 8:123200–123214, 2020.

[5] S. Banerjee and B. Pal. On the enumeration of maximal (δ,γ)-cliques of a temporal network. In *Proceedings of the ACM India Joint International Conference on Data Science and Management of Data*, pages 112–120, 2019.

[6] S. Banerjee and B. Pal. A two-phase approach for enumeration of maximal (δ,γ)-cliques of a temporal network. In *Database and Expert Systems Applications: 32nd International Conference, DEXA 2021, Virtual Event, September 27–30, 2021, Proceedings, Part II 32*, pages 346–357. Springer, 2021.

[7] S. Banerjee and B. Pal. An efficient updation approach for enumerating maximal (δ, γ)-cliques of a temporal network. *Journal of Complex Networks*, 10(5):cnac027, 2022.

[8] S. Banerjee and B. Pal. A two-phase approach for enumeration of maximal (δ,γ)-cliques of a temporal network. *Social Network Analysis and Mining*, 14(1):54, 2024.

[9] A.-L. Barabasi. The origin of bursts and heavy tails in human dynamics. *Nature*, 435(7039):207–211, 2005.

[10] A.-L. Barabási and R. Albert. Emergence of scaling in random networks. *Science*, 286(5439):509–512, 1999.

[11] J. Baras and G. Theodorakopoulos. *Path Problems in Networks*. Springer Nature, 2022.

[12] M. Barbouch, F. W. Takes, and S. Verberne. Combining language models and network features for relevance-based tweet classification. In *International Conference on Social Informatics*, pages 15–27. Springer, 2020.

[13] A. Bavelas. Communication patterns in task-oriented groups. *The Journal of the Acoustical Society of America*, 22(6):725–730, 1950.

[14] A. R. Benson, D. F. Gleich, and J. Leskovec. Higher-order organization of complex networks. *Science*, 353(6295):163–166, 2016.

[15] G. Bianconi. *Higher-order Networks*. Cambridge University Press, 2021.

[16] C. Bick, E. Gross, H. A. Harrington, and M. T. Schaub. What are higher-order networks? *SIAM review*, 65(3):686–731, 2023.

[17] H. Boekhout, I. van der Weijden, and L. Waltman. Gender differences in scientific careers: A large-scale bibliometric analysis. *arXiv preprint arXiv:2106.12624*, 2021.

[18] H. D. Boekhout. Datasets of (aggregated) information on scientific teams supporting results from "freshness, persistence and success of scientific teams", 2025. https://doi.org/10.5281/zenodo.15967685.

[19] H. D. Boekhout, A. A. Blokland, and F. W. Takes. A large-scale longitudinal structured dataset of the dark web cryptomarket Evolution (2014–2015). *arXiv preprint arXiv:2311.11878*, 2023.

[20] H. D. Boekhout, A. A. Blokland, and F. W. Takes. A large-scale longitudinal structured dataset of the dark web cryptomarket Evolution (2014—2015), 2023. https://doi.org/10.5281/zenodo.10171217.

[21] H. D. Boekhout, A. A. Blokland, and F. W. Takes. Early warning signals for predicting cryptomarket vendor success using dark net forum networks. *Scientific Reports*, 14(1):16336, 2024.

[22] H. D. Boekhout, E. M. Heemskerk, N. Pisani, and F. W. Takes. Freshness, persistence and success of scientific teams. *arXiv preprint arXiv:2507.12255*, 2025.

[23] H. D. Boekhout, E. M. Heemskerk, and F. W. Takes. Evolution of the world stage of global science from a scientific city network perspective. In *Complex Networks & Their Applications X*, pages 142–154. Springer International Publishing, 2022.

[24] H. D. Boekhout, W. A. Kosters, and F. W. Takes. Counting multilayer temporal motifs in complex networks. In *International Conference on Complex Networks and their Applications*, pages 565–577. Springer, 2018.

[25] H. D. Boekhout, W. A. Kosters, and F. W. Takes. Efficiently counting complex multilayer temporal motifs in large-scale networks. *Computational Social Networks*, 6(1):8, 2019.

[26] H. D. Boekhout and F. W. Takes. Fast maximal clique enumeration in weighted temporal networks. *Social Network Analysis and Mining*, 16(1):10, 2026.

[27] H. D. Boekhout, V. A. Traag, and F. W. Takes. Investigating scientific mobility in co-authorship networks using multilayer temporal motifs. *Network Science*, 9(3):354–386, 2021.

[28] T. Bol, M. De Vaan, and A. van de Rijt. The matthew effect in science funding. *Proceedings of the National Academy of Sciences*, 115(19):4887–4890, 2018.

[29] P. Boldi and S. Vigna. Axioms for centrality. *Internet Mathematics*, 10(3–4):222–262, 2014.

[30] T. M. Booij, T. Verburgh, F. Falconieri, and R. S. van Wegberg. Get rich or keep tryin' trajectories in dark net market vendor careers. In *2021 IEEE European Symposium on Security and Privacy Workshops (EuroS&PW)*, pages 202–212. IEEE, 2021.

[31] L. Bornmann and F. de Moya-Anegón. Spatial bibliometrics on the city level. *Journal of Information Science*, 45(3):416–425, 2019.

[32] D. Bouyssou and T. Marchant. Ranking authors using fractional counting of citations: An axiomatic approach. *Journal of Informetrics*, 10(1):183–199, 2016.

[33] U. Brandes. A faster algorithm for betweenness centrality. *Journal of Mathematical Sociology*, 25(2):163–177, 2001.

[34] G. Branwen, N. Christin, D. Décary-Hétu, R. M. Andersen, StExo, E. Presidente, Anonymous, D. Lau, D. K. Sohhlz, V. Cakic, V. Buskirk, Whom, M. McKenna, and S. Goode. Dark net market archives, 2011–2015. `https://www.gwern.net/DNM-archives`, July 2015. Accessed: 2021-07-22.

[35] C. Bron and J. Kerbosch. Algorithm 457: Finding all cliques of an undirected graph. *Communications of the ACM*, 16(9):575–577, 1973.

[36] Y. Bu, Y. Ding, X. Liang, and D. S. Murray. Understanding persistent scientific collaboration. *Journal of the Association for Information Science and Technology*, 69(3):438–448, 2018.

[37] E. Caron and N. J. van Eck. Large scale author name disambiguation using rule-based scoring and clustering. In *Proceedings of the 19th International Conference on Science and Technology Indicators*, pages 79–86. CWTS-Leiden University, 2014.

[38] R. Catini, D. Karamshuk, O. Penner, and M. Riccaboni. Identifying geographic clusters: A network analytic approach. *Research Policy*, 44(9):1749–1762, 2015.

[39] L. Cavallaro, A. Ficara, P. De Meo, G. Fiumara, S. Catanese, O. Bagdasar, W. Song, and A. Liotta. Disrupting resilient criminal networks through data analysis: the case of sicilian mafia. *PLOS ONE*, 15(8):e0236476, 2020.

[40] Chainalysis. The Chainalysis 2021 crypto crime report. `https://go.chainalysis.com/2021-Crypto-Crime-Report.html`, 2021.

[41] D. Chen, L. Lü, M.-S. Shang, Y.-C. Zhang, and T. Zhou. Identifying influential nodes in complex networks. *Physica A: Statistical Mechanics and its Applications*, 391(4):1777–1787, 2012.

[42] S. C. Chen and G. S. Dhillon. Interpreting dimensions of consumer trust in e-commerce. *Information Technology and Management*, 4:303–318, 2003.

[43] Z. Chen, L. Yuan, X. Lin, L. Qin, and J. Yang. Efficient maximal balanced clique enumeration in signed networks. In *Proceedings of the Web Conference 2020*, pages 339–349, 2020.

[44] J. Cheng, L. Zhu, Y. Ke, and S. Chu. Fast algorithms for maximal clique enumeration with limited memory. In *Proceedings of the 18th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pages 1240–1248, 2012.

[45] P. Christen. *The Data Matching Process*. Springer, 2012.

[46] N. Christin. Traveling the silk road: A measurement analysis of a large anonymous online marketplace. In *Proceedings of the 22nd International Conference on World Wide Web*, pages 213–224, 2013.

[47] Web of Science categories. `https://images.webofknowledge.com/images/help/WOS/hp_subject_category_terms_tasca.html`. Accessed: Aug. 31, 2021.

[48] A. Conte, D. Firmani, C. Mordente, M. Patrignani, and R. Torlone. Fast enumeration of large k-plexes. In *Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pages 115–124, 2017.

[49] B. Cronin and L. Meho. Using the h-index to rank influential information scientists. *Journal of the American Society for Information Science and Technology*, 57(9):1275–1278, 2006.

[50] G. Csomós. On the challenges ahead of spatial scientometrics focusing on the city level. *Aslib Journal of Information Management*, 72(1):67–87, 2019.

[51] G. Csárdi and T. Nepusz. The igraph software package for complex network research. *InterJournal Complex Systems*, 1695(5):1–9, 2006.

[52] J. N. Cummings and S. Kiesler. Collaborative research across disciplinary and organizational boundaries. *Social studies of science*, 35(5):703–722, 2005.

[53] J. N. Cummings and S. Kiesler. Coordination costs and project outcomes in multi-university collaborations. *Research policy*, 36(10):1620–1634, 2007.

[54] Q. Dai, R.-H. Li, M. Liao, and G. Wang. Maximal defective clique enumeration. *Proceedings of the ACM on Management of Data*, 1(1):1–26, 2023.

[55] A. Das, S.-V. Sanei-Mehri, and S. Tirthapura. Shared-memory parallel maximal clique enumeration. In *2018 IEEE 25th International Conference on High Performance Computing (HiPC)*, pages 62–71. IEEE, 2018.

[56] DEEPDOTWEB. Evolution market background: Carding forums, Ponzi schemes & LE. *DEEP.DOT.WEB article* `https://www.gwern.net/docs/darknet-markets/evolution/2015-03-18-the_avid-evolutionmarketbackground-cardingforumsponzischemesle.html`, March 2015. Accessed: March 2023.

[57] Y. Ding, E. Yan, A. Frazho, and J. Caverlee. PageRank for ranking authors in co-citation networks. *Journal of the American Society for Information Science and Technology*, 60(11):2229–2243, 2009.

[58] R. Dingledine, N. Mathewson, and P. Syverson. Tor: The second-generation onion router. Technical report, Naval Research Lab, Washington DC, 2004.

[59] L. Dioşan, A. Andreica, et al. Network motifs: A key variable in the equation of dynamic flow between macro and micro layers in complex networks. *Knowledge-Based Systems*, 213:106648, 2021.

[60] S. W. Duxbury and D. L. Haynie. The network structure of opioid distribution on a darknet cryptomarket. *Journal of Quantitative Criminology*, 34:921–941, 2018.

[61] J. D. Eblen, C. A. Phillips, G. L. Rogers, and M. A. Langston. The maximum clique enumeration problem: Algorithms, applications, and implementations. In *BMC Bioinformatics*, volume 13, pages 1–11. Springer, 2012.

[62] M. Egan. What is the dark web, what's on it & how to access it. *Tech Advisor how-to article*, `https://www.techadvisor.com/article/727316/what-is-the-dark-web-whats-on-it-how-to-access-it.html`, 2019. Accessed: March 2023.

[63] D. Eppstein, M. Löffler, and D. Strash. Listing all maximal cliques in sparse graphs in near-optimal time. In *Algorithms and Computation: 21st International Symposium, ISAAC 2010, Proceedings, Part I 21*, pages 403–414. Springer, 2010.

[64] M. Familmaleki, A. Aghighi, and K. Hamidi. Analyzing the influence of sales promotion on customer purchasing behavior. *International Journal of Economics & Management Sciences*, 4(4):1–6, 2015.

[65] W. Fan and F. Geerts. *Foundations of data quality management*. Springer Nature, 2022.

[66] A. M. Fonhof, M. van der Bruggen, and F. W. Takes. Characterizing key players in child exploitation networks on the dark net. In *International Conference on Complex Networks and their Applications*, pages 412–423. Springer, 2018.

[67] S. Fortunato, C. T. Bergstrom, K. Börner, J. A. Evans, D. Helbing, S. Milojević, A. M. Petersen, F. Radicchi, R. Sinatra, B. Uzzi, et al. Science of science. *Science*, 359(6379):eaao0185, 2018.

[68] L. C. Freeman. A set of measures of centrality based on betweenness. *Sociometry*, 40(1):35–41, 1977.

[69] R. Guimera, B. Uzzi, J. Spiro, and L. A. N. Amaral. Team assembly mechanisms determine collaboration network structure and team performance. *Science*, 308(5722):697–702, 2005.

[70] A.-S. Himmel, H. Molter, R. Niedermeier, and M. Sorge. Adapting the Bron–Kerbosch algorithm for enumerating maximal cliques in temporal graphs. *Social Network Analysis and Mining*, 7:1–16, 2017.

[71] N. Hiramoto and Y. Tsuchiya. Are illicit drugs a driving force for cryptomarket leadership? *Journal of Drug Issues*, 53(3):451–474, 2023.

[72] P. Holme and J. Saramäki. *Temporal Network Theory*, volume 2. Springer, 2019.

[73] B. Hou, Z. Wang, Q. Chen, B. Suo, C. Fang, Z. Li, and Z. G. Ives. Efficient maximal clique enumeration over graph data. *Data Science and Engineering*, 1:219–230, 2016.

[74] D. Hsiehchen, M. Espinoza, and A. Hsieh. Multinational teams and diseconomies of scale in collaborative research. *Science advances*, 1(8):e1500211, 2015.

[75] R. Huang and E. Sarigöllü. How brand awareness relates to market outcome, brand equity, and the marketing mix. *Journal of Business Research*, 65(1):92–99, 2012.

[76] L. Isella, J. Stehlé, A. Barrat, C. Cattuto, J. Pinton, and W. Van den Broeck. What's in a crowd? Analysis of face-to-face behavioral networks. *Journal of Theoretical Biology*, 271(1):166–180, 2011.

[77] H. Ito and K. Iwama. Enumeration of isolated cliques and pseudo-cliques. *ACM Transactions on Algorithms*, 5(4):1–21, 2009.

[78] W. Jo, S. H. Jang, and E. K. Shin. Stage distinctive communication networks of the online breast cancer community. *Scientific Reports*, 13(1):1726, 2023.

[79] G. Kamphausen and B. Werse. Digital figurations in the online trade of illicit drugs: a qualitative content analysis of darknet forums. *International Journal of Drug Policy*, 73:281–287, 2019.

[80] D. Kempe, J. Kleinberg, and A. Kumar. Connectivity and inference problems for temporal networks. In *Proceedings of the Thirty-second Annual ACM Symposium on Theory of Computing*, pages 504–513, 2000.

[81] M. Kivelä, A. Arenas, M. Barthelemy, J. P. Gleeson, Y. Moreno, and M. A. Porter. Multilayer networks. *Journal of Complex Networks*, 2(3):203–271, 2014.

[82] B. Klimt and Y. Yang. The enron corpus: A new dataset for email classification research. In *Machine Learning: ECML 2004, Lecture Notes in Computer Science*, volume 3201, pages 217–226, Berlin, Heidelberg, 2004. Springer.

[83] I. Konstas, V. Stathopoulos, and J. M. Jose. On social networks and collaborative recommendation. In *Proceedings of the 32nd International ACM SIGIR Conference on Research and Development in Information Retrieval*, pages 195–202. ACM, 2009.

[84] F. Kose, W. Weckwerth, T. Linke, and O. Fiehn. Visualizing plant metabolomic correlation networks using clique–metabolite matrices. *Bioinformatics*, 17(12):1198–1208, 2001.

[85] S. Kumar, W. L. Hamilton, J. Leskovec, and D. Jurafsky. Community interaction and conflict on the web. In *Proceedings of the 2018 World Wide Web Conference on World Wide Web*, pages 933–943, 2018.

[86] S. Kumar, B. Hooi, D. Makhija, M. Kumar, C. Faloutsos, and V. Subrahmanian. REV2: Fraudulent user prediction in rating platforms. In *Proceedings of the Eleventh ACM International Conference on Web Search and Data Mining*, pages 333–341, 2018.

[87] S. Kumar, F. Spezzano, V. Subrahmanian, and C. Faloutsos. Edge weight prediction in weighted signed networks. In *2016 IEEE 16th International Conference on Data Mining (ICDM)*, pages 221–230. IEEE, 2016.

[88] J. Kunegis. KONECT: The Koblenz Network Collection. In *Proceedings of the 22nd International Conference on World Wide Web*, WWW '13 Companion, page 1343–1350, New York, NY, USA, 2013. ACM.

[89] A. N. Langville and C. D. Meyer. *Who's# 1? The science of rating and ranking*. Princeton University Press, 2012.

[90] V. Larivière, Y. Gingras, C. R. Sugimoto, and A. Tsou. Team size matters: Collaboration and scientific impact since 1900. *Journal of the Association for Information Science and Technology*, 66(7):1323–1332, 2015.

[91] J.-H. Lee, N. Sato, K. Yano, and Y. Miyake. Universal association between depressive symptoms and social-network structures in the workplace. *Scientific Reports*, 12(1):10170, 2022.

[92] R. Leon Garcia. Enumerating cliques and their relaxations: Sequential and parallel algorithms. *PhD Thesis, Università degli Studi di Roma "La Sapienza"*, 2019.

[93] J. Leskovec and A. Krevl. SNAP Datasets: Stanford large network dataset collection. `http://snap.stanford.edu/data`, June 2014. Accessed: 2024-06-10.

[94] M.-X. Li, Z.-Q. Jiang, W.-J. Xie, S. Miccichè, M. Tumminello, W.-X. Zhou, and R. N. Mantegna. A comparative analysis of the statistical properties of large mobile phone calling networks. *Scientific Reports*, 4(1):1–12, 2014.

[95] Z. Li, X. Du, X. Liao, X. Jiang, and T. Champagne-Langabeer. Demystifying the dark web opioid trade: Content analysis on anonymous market listings and forum posts. *Journal of Medical Internet Research*, 23(2):e24486, 2021.

[96] M. Liu, A. Jaiswal, Y. Bu, C. Min, S. Yang, Z. Liu, D. Acuña, and Y. Ding. Team formation and team impact: The balance between team freshness and repeat collaboration. *Journal of Informetrics*, 16(4):101337, 2022.

[97] X. Liu, E. Patacchini, Y. Zenou, and L.-F. Lee. Criminal networks: Who is the key player? *FEEM Working Paper*, 39, 2012.

[98] X. Lu and C. Brelsford. Network structure and community evolution on Twitter: Human behavior change in response to the 2011 Japanese earthquake and tsunami. *Scientific Reports*, 4(1):6773, 2014.

[99] Z. Ma, L. Wang, J. Yang, X. Zhou, K. Li, and C. Gao. Maximal η-clique maintenance over uncertain graph streams. *Information Sciences*, page 122318, 2025.

[100] M. Maisonobe, D. Eckert, M. Grossetti, L. Jégou, and B. Milard. The world network of scientific collaborations between cities: Domestic or international dynamics? *Journal of Informetrics*, 10(4):1025–1036, 2016.

[101] M. Maisonobe, L. Jégou, and D. Eckert. Delineating urban agglomerations across the world: A dataset for studying the spatial distribution of academic research at city level. *Cybergeo: European Journal of Geography*, 2018. Document 871.

[102] G. Manoussakis. A new decomposition technique for maximal clique enumeration for sparse graphs. *Theoretical Computer Science*, 770:25–33, 2019.

[103] J. Martin. Lost on the Silk Road: Online drug distribution and the 'cryptomarket'. *Criminology & Criminal Justice*, 14(3):351–367, 2014.

[104] R. K. Merton. The matthew effect in science: The reward and communication systems of science are considered. *Science*, 159(3810):56–63, 1968.

[105] K. Moeller, R. Munksgaard, and J. Demant. Flow my FE the vendor said: Exploring violent and fraudulent resource exchanges on cryptomarkets for illicit drugs. *American Behavioral Scientist*, 61(11):1427–1450, 2017.

[106] A. P. Mukherjee, P. Xu, and S. Tirthapura. Enumeration of maximal cliques from an uncertain graph. *IEEE Transactions on Knowledge and Data Engineering*, 29(3):543–555, 2016.

[107] S. Mukherjee, D. M. Romero, B. Jones, and B. Uzzi. The nearly universal link between the age of past knowledge and tomorrow's breakthroughs in science and technology: The hotspot. *Science Advances*, 3(4):e1601315, 2017.

[108] R. Munksgaard and J. Demant. Mixing politics and crime–the prevalence and decline of political discourse on the cryptomarket. *International Journal of Drug Policy*, 35:77–83, 2016.

[109] M. Nadini, A. Bracci, A. ElBahrawy, P. Gradwell, A. Teytelboym, and A. Baronchelli. Emergence and structure of decentralised trade networks around dark web marketplaces. *Scientific Reports*, 12(1):1–9, 2022.

[110] S. Nakamoto. Bitcoin: A peer-to-peer electronic cash system. *Satoshi Nakamoto [Online]*, 2008. Available: `https://www.bitcoinpaper.info/bitcoinpaper-html/`.

[111] R. Narang, S. Misra, and R. Goyal. An empirical study on the role of macro-meso-micro measures in citation networks. In *Computational Science and Its Applications–ICCSA 2019: 19th International Conference 2019, Proceedings, Part IV 19*, pages 340–356. Springer, 2019.

[112] L. Norbutas, S. Ruiter, and R. Corten. Reputation transferability across contexts: Maintaining cooperation among anonymous cryptomarket actors when moving between markets. *International Journal of Drug Policy*, 76:102635, 2020.

[113] E. S. Olivares, H. D. Boekhout, A. Saxena, and F. W. Takes. A framework for empirically evaluating pretrained link prediction models. In *International Conference on Complex Networks and Their Applications*, pages 150–161. Springer, 2023.

[114] L. Page, S. Brin, R. Motwani, and T. Winograd. The PageRank citation ranking: Bringing order to the web. Technical report, Stanford InfoLab, 1999.

[115] P. Panzarasa, T. Opsahl, and K. M. Carley. Patterns and dynamics of users' behavior and interaction: Network analysis of an online community. *Journal of the American Society for Information Science and Technology*, 60(5):911–932, 2009.

[116] N. Pisani, H. D. Boekhout, E. M. Heemskerk, and F. W. Takes. China's rise as global scientific powerhouse: A trajectory of international collaboration and specialization in high-impact research. *Research Policy*, 54(8):105288, 2025.

[117] O. Poquet, L. Tupikina, and M. Santolini. Are forum networks social networks? A methodological perspective. In *Proceedings of the Tenth International Conference on Learning Analytics & Knowledge*, pages 366–375, 2020.

[118] R. Pranckutė. Web of Science (WoS) and Scopus: The titans of bibliographic information in today's academic world. *Publications*, 9(1):12, 2021.

[119] L. Putman, H. D. Boekhout, and F. W. Takes. Fast incremental computation of harmonic closeness centrality in directed weighted networks. In *Proceedings of the 2019 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining*, pages 1018–1025, 2019.

[120] H. Qin, R.-H. Li, G. Wang, L. Qin, Y. Cheng, and Y. Yuan. Mining periodic cliques in temporal networks. In *2019 IEEE 35th International Conference on Data Engineering (ICDE)*, pages 1130–1141. IEEE, 2019.

[121] T. Reisch, G. Heiler, C. Diem, P. Klimek, and S. Thurner. Monitoring supply networks from mobile phone data for estimating the systemic risk of an economy. *Scientific Reports*, 12(1):13347, 2022.

[122] E. Renyi. On random graph. *Publicationes Mathematicate*, 6:290–297, 1959.

[123] D. Rhumorbarbe, L. Staehli, J. Broséus, Q. Rossy, and P. Esseiva. Buying drugs on a darknet market: A better deal? Studying the online illicit drug market through the analysis of digital, physical and chemical data. *Forensic Science International*, 267:173–182, 2016.

[124] Y. Rochat. Closeness centrality extended to unconnected graphs: The harmonic centrality index. Technical report, ASNA, 2009.

[125] B. Ruhnau. Eigenvector-centrality—a node-centrality? *Social Networks*, 22(4):357–365, 2000.

[126] M. C. Schmidt, N. F. Samatova, K. Thomas, and B.-H. Park. A scalable, parallel algorithm for maximal clique enumeration. *Journal of Parallel and Distributed Computing*, 69(4):417–428, 2009.

[127] H. Shirado and N. A. Christakis. Locally noisy autonomous agents improve global human coordination in network experiments. *Nature*, 545(7654):370–374, 2017.

[128] P. Shortis, J. Aldridge, and J. Monica. Drug cryptomarket futures: Structure, function and evolution in response to law enforcement actions. In *Research Handbook on International Drug Policy*, pages 355–380. Edward Elgar Publishing, 2020.

[129] G. Simmel. *Soziologie*. Duncker & Humblot, 1908.

[130] G. Soda, P. V. Mannucci, and R. S. Burt. Networks, creativity, and time: Staying creative through brokerage and network rejuvenation. *Academy of Management Journal*, 64(4):1164–1190, 2021.

[131] F. W. Takes and E. M. Heemskerk. Centrality in the global network of corporate control. *Social Network Analysis and Mining*, 6(1):1–18, 2016.

[132] S. Tannenbaum and E. Salas. *Teams that work: The seven drivers of team effectiveness*. Oxford University Press, 2020.

[133] E. Tomita, A. Tanaka, and H. Takahashi. The worst-case time complexity for generating all maximal cliques and computational experiments. *Theoretical Computer Science*, 363(1):28–42, 2006.

[134] T. Uno. An efficient algorithm for solving pseudo clique enumeration problem. *Algorithmica*, 56:3–16, 2010.

[135] B. Uzzi, S. Mukherjee, M. Stringer, and B. Jones. Atypical combinations and scientific impact. *Science*, 342(6157):468–472, 2013.

[136] J. E. van Engelen, H. D. Boekhout, and F. W. Takes. Explainable and efficient link prediction in real-world network data. In *International Symposium on Intelligent Data Analysis*, pages 295–307. Springer, 2016.

[137] M. C. Van Hout and E. Hearne. New psychoactive substances (NPS) on cryptomarket fora: An exploratory study of characteristics of forum activity between NPS buyers and vendors. *International Journal of Drug Policy*, 40:102–110, 2017.

[138] A. Vespignani. Twenty years of network science. *Nature*, 558:528–529, 2018.

[139] J. Viard, M. Latapy, and C. Magnien. Revealing contact patterns among high-school students using maximal cliques in link streams. In *Proceedings of the*

2015 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining 2015, pages 1517–1522, 2015.

[140] T. Viard, M. Latapy, and C. Magnien. Computing maximal cliques in link streams. *Theoretical Computer Science*, 609:245–252, 2016.

[141] T. Viard, C. Magnien, and M. Latapy. Enumerating maximal cliques in link streams with durations. *Information Processing Letters*, 133:44–48, 2018.

[142] B. Viswanath, A. Mislove, M. Cha, and K. P. Gummadi. On the evolution of user interaction in facebook. In *Proceedings of the 2nd ACM Workshop on Online Social Networks*, pages 37–42, 2009.

[143] L. Waltman, C. Calero-Medina, J. Kosten, E. C. Noyons, R. J. Tijssen, N. J. van Eck, T. N. van Leeuwen, A. F. van Raan, M. S. Visser, and P. Wouters. The Leiden Ranking 2011/2012. *Journal of the American Society for Information Science and Technology*, 63(12):2419–2432, 2012.

[144] L. Waltman, C. Calero-Medina, J. Kosten, E. C. Noyons, R. J. Tijssen, N. J. van Eck, T. N. van Leeuwen, A. F. van Raan, M. S. Visser, and P. Wouters. The Leiden Ranking 2011/2012: Data collection, indicators, and interpretation. *Journal of the American Society for Information Science and Technology*, 63(12):2419–2432, 2012.

[145] L. Waltman, N. J. van Eck, T. N. van Leeuwen, M. S. Visser, and A. F. van Raan. Towards a new crown indicator: An empirical analysis. *Scientometrics*, 87(3):467–481, 2011.

[146] D. Wang and A.-L. Barabási. *The Science of Science*. Cambridge University Press, 2021.

[147] J. Wang, J. Yang, Z. Gu, D. Ouyang, Z. Tian, and X. Lin. Efficient maximal biclique enumeration on large signed bipartite graphs. *IEEE Transactions on Knowledge and Data Engineering*, 36(9):4618–4631, 2024.

[148] J. Wang, J. Yang, Z. Ma, C. Zhang, S. Yang, and W. Zhang. Efficient maximal biclique enumeration on large uncertain bipartite graphs. *IEEE Transactions on Knowledge and Data Engineering*, 35(12):12634–12648, 2023.

[149] X. Wang, J. Li, E. Srivatsavaya, and S. Rajtmajer. Evidence of inter-state coordination amongst state-backed information operations. *Scientific Reports*, 13(1):7716, 2023.

[150] I. Wanzenböck, R. Lata, and D. Ince. Proposal success in horizon 2020: A study of the influence of consortium characteristics. *Quantitative Science Studies*, 1(3):1136–1158, 2020.

[151] D. J. Watts and S. H. Strogatz. Collective dynamics of 'small-world' networks. *Nature*, 393(6684):440–442, 1998.

[152] C. Wilson, B. Boe, A. Sala, K. P. Puttaswamy, and B. Y. Zhao. User interactions in social networks and their implications. In *Proceedings of the 4th ACM European conference on Computer systems*, pages 205–218, 2009.

[153] World university ranking. `https://www.timeshighereducation.com/world-university-rankings`. Accessed: Sep. 6, 2021.

[154] B. Wu, S. Yang, H. Zhao, and B. Wang. A distributed algorithm to enumerate all maximal cliques in mapreduce. In *2009 Fourth International Conference on Frontier of Computer Science and Technology*, pages 45–51. IEEE, 2009.

[155] S. Wuchty, B. F. Jones, and B. Uzzi. The increasing dominance of teams in production of knowledge. *Science*, 316(5827):1036–1039, 2007.

[156] A. Zeng, Y. Fan, Z. Di, Y. Wang, and S. Havlin. Fresh teams are associated with original and multidisciplinary research. *Nature human behaviour*, 5(10):1314–1322, 2021.

[157] C. Zhang, Y. Zhang, W. Zhang, L. Qin, and J. Yang. Efficient maximal spatial clique enumeration. In *2019 IEEE 35th International Conference on Data Engineering (ICDE)*, pages 878–889. IEEE, 2019.

[158] K. Zhu, G. Fletcher, N. Yakovets, O. Papapetrou, and Y. Wu. Scalable temporal clique enumeration. In *Proceedings of the 16th International Symposium on Spatial and Temporal Databases*, pages 120–129, 2019.