Analysis of Complex Customer Networks: A Real-World Banking Example

Karmela Ljubičić*, Andro Merćep†, Zvonko Kostanjčar‡
* Data Warehouse & Business Intelligence Department, Privredna banka Zagreb, 10000 Zagreb, Croatia
†‡ Laboratory for Financial and Risk Analytics, Faculty of Electrical Engineering and Computing, University of Zagreb, 10000 Zagreb, Croatia
karmela.ljubicic@{pbz.hr, fer.hr}*, andro.mercep@fer.hr†, zvonko.kostanjcar@fer.hr‡

***Abstract*—Complex customer networks are an important tool for better understanding of customer behaviour and customers' mutual interdependence. Hence, they have proven to be insightful in numerous customer-centric domains such as recommender systems and customer relationship management. Although previous research in this field is predominantly associated with telecommunication and e-commerce sectors, similar construction principles could be transferred to the financial sector, e.g. a customer network could be created based on transaction history or domain-specific relations such as loan debtor–guarantor or loan debtor–co-debtor. However, this research path is still underdeveloped. In this paper, we analyse real-world complex customer networks of a Croatian bank. Numerous graph metrics were calculated, and the obtained results prove that these networks are non-random. We also show that these networks are scale-free and exhibit the small-world effect based on graph properties like the node degree distribution and the average shortest path.**

Keywords*—*complex networks, graph metrics, social network analysis, transaction networks, banking

I. INTRODUCTION

In customer-driven industries and in current big-data era, a large quantity of customer data is available to service providers. With a higher customer satisfaction as the final goal, more and more companies are turning to complex network analysis to deepen their knowledge about their customers, their mutual interactions and possible ways to encourage them to participate in actions desirable to the companies, e.g. buying a certain product.

Improved insight in customer behaviour is facilitated through complex network analysis as a part of graph theory. Although it stems from discrete mathematics, a great number of various scientific and business fields have leveraged this research in recent years [1]. One of them is social network analysis (SNA), which flourished in the last decade with emergence of large online social networks. SNA is a research area focusing on relationships among social entities, as well as the implications of those relationships [2]. Patterns in social groups represented by networks can be exploited in many science and business domains by examining network statistical properties such as degree distributions and path lengths. These topological characteristics can help predict the performance of different processes taking place in the network (e.g. spreading of influence). Excellent overviews of network theory can be found in [3] and [4], while detailed examples of massive social network analysis are provided in [5]–[8]. Social concepts like Word-Of-Mouth and influence spreading can be observed not only in explicitly given social groups (e.g. friends on Facebook), but also in networks where the social connections between people are implicit [9]–[11]. For example, if two customers of a telecommunication company communicate often, it is reasonable to assume that they know each other in real life (although the type of their real social connection remains unknown). Therefore, service providers use complex network analysis and SNA principles to build customer-centric networks in which the service customers and their connections are modeled as nodes and edges, respectively.

Empirically observed graph metrics of different customer networks were presented in literature, while the naming of the graphs is dependent on the application domain. A considerable number of studies focused on improving recommender systems in the field of e-commerce with graph-derived knowledge. Improved network marketing was achieved with complex network analysis in [12]–[15], while using the terms: consumer network, recommendation network, commerce network or customer-product network. In the telecommunication and financial sectors, networks are often named by the type of data source used for network construction. For example, Call Detail Records (CDR) data hold information about the caller, callee, call date, and call duration, and based on the given information, a CDR/call/SMS graph can be created and systematic analysis of these structures was offered in [16]–[18]. In [19], a customer network was created using real-world CDR data of a mobile phone provider to find influential customers that would spread viral marketing campaigns most efficiently.

Similar construction principles could be transferred to the financial sector, in which transaction history is a main data source for various use cases. To the best of our knowledge, there are not many studies in this field. With the information about the payer, payee, payment date, and amount, a transaction network can easily be created. Recently, a couple of studies analysed the cryptocurrency transaction networks [20], [21]. More traditional use cases in banking, for which transaction networks were created and studied, are anomaly detection [22], credit risk assess-

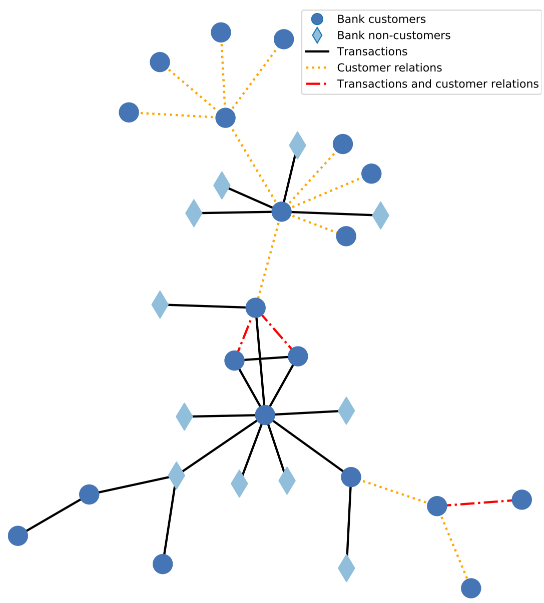

Fig. 1: Illustration of a bank customer graph

ment [23], identifying influential clients [24] and cash-flow prediction [25].

Therefore, our goal is to further expand the knowledge about the topology of bank customer networks and their properties. In this paper we offer a method for bank customer network construction and a detailed analysis of the resulting complex structure. We show that the constructed bank customer networks are non-random. Moreover, the present graph characteristics confirm that the customer networks correspond to a great number of real-world networks in terms of the scale-free and the small-world properties.

II. BANK CUSTOMER NETWORK

Real social ties in financial customer networks are not explicit, in contrast to those present in some online social networks (e.g. online friendships). Since the bank has limited information about customers' real social interactions, their proxy must be derived from the available customer data. In this section, we present the method used for bank customer network construction.

A. Definitions

A network (also called a graph) $\mathcal{G} = (\mathcal{N}, \mathcal{E})$ is a construction of nodes $\mathcal{N}$ and edges $\mathcal{E}$ connecting them. In the context of complex customer networks, every node $n \in \mathcal{N}$ represents a single customer and every edge $e \in \mathcal{E}$ represents a derived social connection between two nodes, i.e. the nodes incident to the edge e. Nodes that are connected with an edge are called adjacent nodes or neighbours. If all nodes are neighbours with each other, the graph is complete (fully connected).

Additionally, edges can be directed and their direction represents the assumed path of information transfer (e.g. from caller to callee). However, if the assumed path goes in both directions, the edges are effectively undirected. Hence, graphs can be defined as either directed or undirected. In addition, every edge can have an associated weight representing the strength of the connection between customers with a numerical value (e.g. an edge between customers that communicate frequently would have a larger weight). In case of weighted edges, the earlier unweighted graph definition is extended with a set of associated edge weights $\mathcal{W}$, i.e. $\mathcal{G} = (\mathcal{N}, \mathcal{E}, \mathcal{W})$.

B. Network construction method

Inspired by the approaches in the literature, we have used both bank customer relations [24] and transaction history [23]–[25] to create complex bank customer networks. From the transaction history data, we filtered the transactions in which at least one bank customer was involved. Although the transaction parties are identified by account, multiple accounts associated with the same bank customer are aggregated into a single unit. However, this is not possible for non-customer accounts because their information is unknown to the bank. Therefore, both single and aggregated customer accounts will be referred to as customers, while non-customer accounts will be referred to as non-customers for brevity.

For every month of construction t, a graph $\mathcal{G}^t$ is created using the available customer data from time period $[t-\tau, t]$. The parameter τ is equal to 12 months and approximates a time span in which individuals interact with the people they are closely connected with in real life.

For every graph $\mathcal{G}^t$, we observe the transaction history and bank customer relations that have happened in $[t-\tau, t]$. Every such connection implies an edge with its incidental nodes, which are added to the graph $\mathcal{G}^t$. In summary, a node is added for every:

- bank customer who had had at least 1 money transaction with a bank customer/non-customer in $[t-\tau, t]$,
- non-customer who had participated in at least 1 money transaction with a bank customer in $[t-\tau, t]$,
- bank customer who has been a part of at least 1 bank customer relation in $[t-\tau, t]$,

while the edges between the nodes are added if a specific connection existed in $[t-\tau, t]$. We should note that we are unable to provide any further details on the data, either customer relations or transaction history, due to privacy concerns.

Presented network construction method is an aggregation process performed on a monthly basis to increase the stability of the resulting graphs. We are modelling only whether two customers have been involved in a money transaction or customer relation, rather than focusing on the strength or direction of the connections, which results in undirected and unweighted customer graphs. An illustration of a constructed bank customer graph is given in Fig. 1. Described construction method was repeated for a 12-month period from recent history, which resulted in 12 monthly aggregated bank customer networks.

III. EMPIRICAL ANALYSIS OF THE BANK CUSTOMER NETWORK

A random graph is a graph in which the edges are randomly distributed [26], but this is not the case in many real-world networks. The differences between random an non-random real networks had recently sparked interest of scientists, because their contrasting features could reveal the ways in which the real-world networks could be exploited for better decision making [3]. Following the suggestion that only topological properties of the giant components should be analysed [4], in the next subsections we present the obtained results for various graph metrics calculated on the monthly giant components[1].

A. Component size analysis

Although all 12 constructed graphs are massive in number of nodes, not all nodes are connected. Moreover, a graph $\mathcal{G}^t$ is composed of a number of components, i.e. sets of nodes for which each pair of nodes is connected by at least one path. For real large-scale graphs, the largest connected component is usually orders of magnitude bigger than the second largest connected component [5], [6], [16], [17], [23], which is also true in the presented bank customer case. In Fig. 2, the distribution of component sizes for $\mathcal{G}^{\text{December}}$ is shown. The outlier on the right side of the figure represents the giant component, with 65.25% of all nodes and 84.33% of all edges. We discovered an interesting finding by analysing the customers and their interconnections in several largest components excluding the giant component in every month. Customers and their mutual interconnections from these structures proved the existence of two distinct social groups in our customer networks: 1. tightly connected closed organisations, and 2. friends and families from rural areas.

On the contrary, the analysis of individual customer features and their connections for the giant component is not feasible, because of its size. To better understand the topology of the giant components, structural properties such as degree distribution, characteristic path length and clustering coefficient must be considered [4], and they are described and analysed in the next section.

B. Degree distribution

An interesting node's property is the degree k, which is equal to the number of node's neighbours, and graphs are often described with the average degree $\bar{k}$, which is listed in Table I. For large-scale networks, the fraction of nodes that have precisely k neighbours, i.e. the degree distribution $P(k)$, has a Poisson distribution. However, many real-world networks have a highly right skewed degree distribution [13], [15], [16], [23]. These findings are in line with the degree distributions of our bank customer networks. The degree distribution of our $\mathcal{G}^{\text{December}}$ is illustrated in Fig. 3 on a log-log scale. It is clear that the degree

[1]Results for all months are nearly identical. In cases where the results for only 1 month are shown, others are omitted for conciseness.

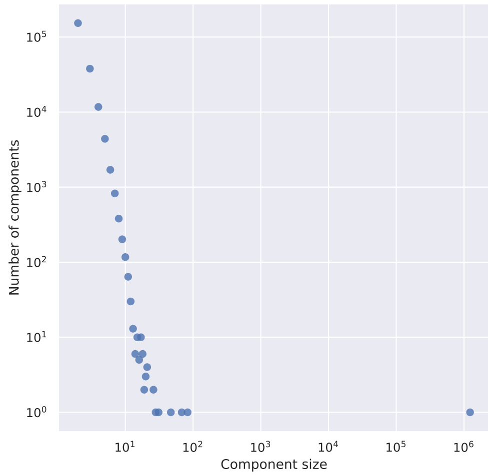

Fig. 2: Distribution of component sizes on a log-log scale for $\mathcal{G}^{\text{December}}$ follows a power-law distribution.

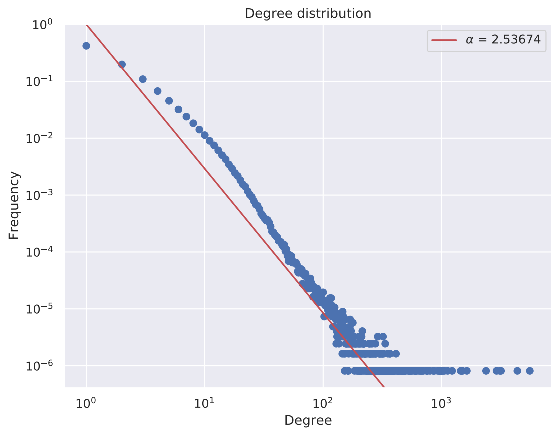

Fig. 3: Distribution of node degree k on a log-log scale for $\mathcal{G}^{\text{December}}$ follows a power-law distribution $P(k) \propto k^{-\alpha}, \alpha = 2.53674$. This is a feature of scale-free networks.

distribution presented as an approximately straight line follows a power-law distribution, mathematically defined with $P(k) \propto k^{-\alpha}, \alpha > 0$.

C. Scale-free networks

Complex networks with a power-law degree distribution, characterised with a long right tail, belong to the family of scale-free networks. Many real world examples of scale-free networks are well described in [15], [16], [19], [23] and [25], with $2 < \alpha < 3$. We report the values of α in Table I and conclude that the obtained degree distributions are in line with similar research.

One of the characteristics of scale-free networks is a heavy tail that indicates the existence of nodes with order of magnitude bigger degrees than the vast majority of nodes in the network. This can also be observed in Fig. 3. The few nodes that stand out in the bottom right corner of the Fig. 3 are called hubs [27]. Analysing the transactions and customer relations that caused such high degrees,

TABLE I: General metrics of the created bank customer networks

Month, t	**Average degree,** $\overline{k}$	**Giant component's average degree** $\overline{k}$	**Scale-free factor,** α **(%)**	**Clustering coefficient,** $C^{(1)}$ **(%)**	**Clustering coefficient,** $C^{(2)}$ **(%)**
January	2.340	3.470	2.550	0.070	7.118
February	2.356	3.482	2.552	0.071	7.105
March	2.378	3.488	2.547	0.073	7.067
April	2.398	3.483	2.546	0.132	7.010
May	2.385	3.458	2.546	1.198	7.039
June	2.412	3.482	2.538	0.940	7.006
July	2.446	3.512	2.535	0.817	6.998
August	2.459	3.523	2.532	0.806	6.994
September	2.493	3.562	2.530	0.824	6.995
October	2.526	3.598	2.536	0.845	7.000
November	2.557	3.626	2.528	0.886	6.997
December	2.623	3.685	2.537	0.883	6.993

we concluded that the hubs are individuals involved in organising humanitarian or commercial activities.

D. Clustering coefficient

Network transitivity, also called clustering, is another frequently reported network feature. Colloquially, it states that the friend of your friend is likely to be your friend. Highly transitive networks are characterised by a large number of triangles – sets of 3 fully connected nodes. This property can be quantified by the clustering coefficient C. Two definitions for C are used in the literature, i.e. the global and the average local clustering coefficient. Inspired by [3], we denote them as $C^{(1)}$ and $C^{(2)}$, respectively. $C^{(1)}$ can be calculated as:

$$C^{(1)} = \frac{3 \cdot \text{total number of triangles in the network}}{\text{total number of connected triples in the network}},$$

where triples are constituted of a node and its 2 neighbours. Effectively, $C^{(1)}$ measures the probability that two randomly selected nodes with a common neighbour are connected themselves. We report our results in Table I. Although the values for $C^{(1)}$ are relatively low, they are several orders of magnitude larger than the clustering coefficient of a random network of the same size, i.e. $C = \mathcal{O}(|\mathcal{N}|^{-1})$ for large $|\mathcal{N}|$. We should note that the $C^{(1)}$ coefficient is rarely reported in the literature regarding the customer network analysis. However, our results are in line with those of many massive real-world networks in Stanford Large Network Dataset Collection [28].

Another widely used definition of the network's clustering coefficient is the average local clustering coefficient: $C^{(2)} = \frac{1}{n}\sum_{i\in\mathcal{N}} C_i$, where C_i is the local clustering coefficient proposed in [29]. It is defined as:

$$C_i = \frac{\text{number of triangles connected to node } i}{\text{number of triples centered on node } i}.$$

In our data set, $C^{(2)}$ is approximately 7% for all months (as seen in Table I). Although this value is not large, it is again several orders of magnitude greater than the clustering coefficient of a random network of the same size [3]. Moreover, it is comparable with results from [6], [19], [20], [24]. In general, $C^{(2)}$ is larger than $C^{(1)}$ in networks with a significant number of low-degree nodes (e.g. scale-free networks) and the differences present in Table I are expected. $C^{(2)}$ tends to be dominated by low-degree nodes, since they are involved in a small number of possible triples (the denominator of C_i). This discrepancy can also be verified in many networks that are available in [28].

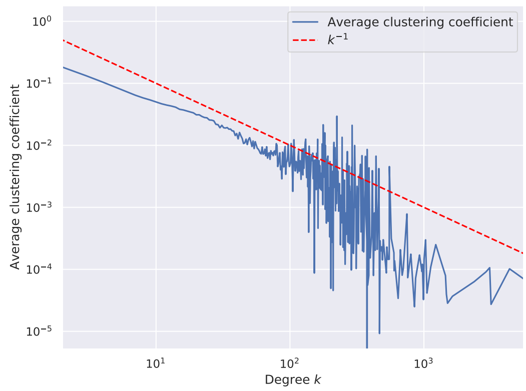

Fig. 4: Dependence of average local clustering coefficient C_k on degree k for $\mathcal{G}^{\text{December}}$.

An interesting property of the scale-free networks was found in [30] and [31], i.e. the dependence of C_i on the degree k_i of a node i is $C_i \approx k_i^{-1}$. This behaviour was noticed empirically in [5], as well as in our study. Fig. 4 shows the average of the local clustering coefficients C_i of nodes with degree k on log-log scale. It is observable that nodes with a lower number of neighbours are more likely to be connected in tighter clusters, in which other nodes are also connected with each other (e.g. friends and family). Another observation that can be made is that after $k = 100$ the average local clustering coefficient starts to oscillate with very low scores. This indicates that bank customers with a higher degree connect to other nodes more indiscriminately (e.g. the hubs from III-C).

E. Maximum clique

In graph theory, a clique [32] is a complete undirected subgraph of at least three nodes. Then, a maximum clique

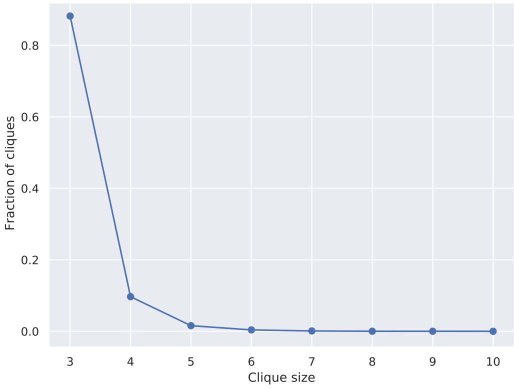

Fig. 5: Power-law distribution of clique sizes for $\mathcal{G}^{\text{December}}$.

TABLE II: Estimated values of characteristic path length L and diameter D

Month, t	**Characteristic path length, L**	**Diameter, D**	**Estimation error, %**
January	6.602	28	0.92
February	6.593	27	1.00
March	6.604	26	1.44
April	6.960	28	2.03
May	8.187	28	5.74
June	8.044	28	4.32
July	7.957	28	5.06
August	7.937	28	5.18
September	7.915	28	5.18
October	7.854	28	5.59
November	7.853	29	6.26
December	7.779	30	5.33

of a network is the clique with the largest number of nodes. The distribution of the clique sizes was studied in [16] and [17]. Our clique size distributions are in line with previous research and we illustrate our results in Fig. 5 for $\mathcal{G}^{\text{December}}$. We found no cliques with more than 10 nodes during the observed 12 month period. The distribution is highly right skewed and follows a power law distribution, with a large number of cliques of size 3. For various service providers, this finding is useful for defining closed user groups [16] and defining specialised incentives for encouraging customer loyalty (e.g. discounts for friends and family) [33], [34].

F. Characteristic path length

Another important structural property of complex networks is the characteristic path length L, calculated as the average shortest path length between any pair of nodes. It takes its minimum value $L = 1$ in case of a complete graph. On the contrary, diameter D is the maximum value of all shortest path lengths found in the network.

To calculate L, path lengths between every pair of nodes must be known. However, brute-forcing this process has a complexity of $O(|\mathcal{N}|^2)$, which is not feasible for very large networks. Therefore, we have estimated the path lengths following the procedure in [35]. The outline of this method is:

1) The procedure starts with a sampled subset of landmark nodes $\mathcal{D} = \{u_1, u_2, ..., u_{|\mathcal{D}|}\}, \mathcal{D} \subset \mathcal{N}$. Landmarks can be selected at random or following one of various proposed selection heuristics (e.g. select top $|\mathcal{D}|$ nodes with the highest degree).
2) The distance between every pair of nodes $d(s,t)$ is calculated using the distances from the nodes s,t to the landmarks $\mathcal{D}$. Using the triangle inequality, $d(s,t)$ is bounded from below by $L = \max_i |d(s,u_i) - d(t,u_i)|$ and bounded from above by $U = \min_j \{d(s,u_j) + d(t,u_j)\}$. Any value in range $[L,U]$ can be used as an estimate $\tilde{d}(s,t)$, but the authors suggest to use the upper bound.

The results in Table II are estimated with $|\mathcal{D}| = 20$ landmarks, chosen with the basic degree heuristic, while using the upper bound as the path length estimation. The authors of this estimation method also proposed the estimation evaluation framework, which we have implemented, and report the estimation error for every month in Table II. The error is equal to the mean discrepancy between real and estimated shortest path lengths for 500 randomly sampled node pairs. From the low estimation error values, it can be concluded that the estimation process was successful.

G. Small-world effect

The first demonstration of the phenomenon known as the small-world effect was carried out by Stanley Milgram in 1967 [36]. His famous experiment with letter passing inside a real social network showed that most people are connected by a short path of people in-between.

Recently, this effect was verified in a large number of social and transaction networks described in [5], [16], [17], [19] and [37], where the values for L are orders of magnitude lower than the network size. The upper-bounded values of L reported in the Table II are in line with the previous research. Implication of these results is that our constructed bank customer networks exhibit the small-world effect. This means that some information (e.g. product recommendation, positive/negative experience with the bank) could spread through the whole bank customer network only in few propagation steps.

IV. Conclusion

Many businesses from customer-focused domains such as e-commerce, telecommunications or banking are constantly exploring new ways to further enhance their relationship with customers and to increase their competitiveness. One option is better utilisation of the available customer data by creating complex customer networks using the knowledge from social network analysis and analysing their topology. In many studies, these structures were proven to be non-random and the discovered properties were leveraged to achieve certain aims (e.g. finding the most influential clients for more efficient marketing campaigns). In this paper, we presented a method for bank customer network construction and the results obtained with complex network analysis. The reported results are analogous to similar social and customer network studies. Large-scale bank customer networks are scale-free networks with a portion of customers representing social hubs. The network's clustering coefficient is orders of magnitude larger when compared to random graphs, while the shortest average path distances imply the existence of the small-world effect. The discrepancy between random and non-random networks is crucial for better utilisation of processes affected by the customer network structural properties, e.g. the diffusion process, and further research is needed to optimally exploit this discrepancy in the banking use case. Finally, an astute reader might notice that our customer graph is a static network, although it could be modelled as a temporal network. For this paper we decided to use time-aggregated networks as simplified static network projections of dynamic networks (stemming from transactions that carry time stamps). Temporal network model would allow us to study many other useful network properties, such as detection of temporal patterns and motifs, identification of temporal node influence, or evolution of the temporal network system. Although this research path surely offers a more detailed insight in customer behaviour, it was outside the scope of this paper and is left for future work.

References

[1] F. Bonchi, C. Castillo, A. Gionis, and A. Jaimes, "Social network analysis and mining for business applications," *ACM Transactions on Intelligent Systems and Technology (TIST)*, vol. 2, no. 3, pp. 1–37, 2011.

[2] S. Wasserman and K. Faust, *Social Network Analysis: Methods and Applications*, ser. Structural Analysis in the Social Sciences. Cambridge University Press, 1994.

[3] M. E. Newman, "The structure and function of complex networks," *SIAM review*, vol. 45, no. 2, pp. 167–256, 2003.

[4] N. Boccara, *Modeling complex systems*. Springer Science & Business Media, 2010.

[5] J. Ugander, B. Karrer, L. Backstrom, and C. Marlow, "The anatomy of the facebook social graph," *arXiv preprint arXiv:1111.4503*, 2011.

[6] S. A. Catanese, P. De Meo, E. Ferrara, G. Fiumara, and A. Provetti, "Crawling facebook for social network analysis purposes," in *Proceedings of the international conference on web intelligence, mining and semantics*, 2011, pp. 1–8.

[7] J. Leskovec and E. Horvitz, "Planetary-scale views on a large instant-messaging network," in *Proceedings of the 17th international conference on World Wide Web*, 2008, pp. 915–924.

[8] A. Java, X. Song, T. Finin, and B. Tseng, "Why we twitter: understanding microblogging usage and communities," in *Proceedings of the 9th WebKDD and 1st SNA-KDD 2007 workshop on Web mining and social network analysis*, 2007, pp. 56–65.

[9] K. Dasgupta, R. Singh, B. Viswanathan, D. Chakraborty, S. Mukherjea, A. A. Nanavati, and A. Joshi, "Social ties and their relevance to churn in mobile telecom networks," in *Proceedings of the 11th international conference on Extending database technology: Advances in database technology*, 2008, pp. 668–677.

[10] Y. Richter, E. Yom-Tov, and N. Slonim, "Predicting customer churn in mobile networks through analysis of social groups," in *Proceedings of the 2010 SIAM international conference on data mining*. SIAM, 2010, pp. 732–741.

[11] W. Verbeke, D. Martens, and B. Baesens, "Social network analysis for customer churn prediction," *Applied Soft Computing*, vol. 14, pp. 431–446, 2014.

[12] S. Hill, F. Provost, and C. Volinsky, "Network-based marketing: Identifying likely adopters via consumer networks," *Statistical Science*, vol. 21, no. 2, pp. 256–276, 2006.

[13] J. Leskovec, L. A. Adamic, and B. A. Huberman, "The dynamics of viral marketing," *ACM Transactions on the Web (TWEB)*, vol. 1, no. 1, pp. 5–es, 2007.

[14] M. Wang, W. Chen, Y. Huang, N. S. Contractor, and Y. Fu, "Modeling customer preferences using multidimensional network analysis in engineering design," *Design Science*, vol. 2, 2016.

[15] Z. Shen and N. Sundaresan, "ebay: an e-commerce marketplace as a complex network," in *Proceedings of the fourth ACM international conference on Web search and data mining*, 2011, pp. 655–664.

[16] A. A. Nanavati, S. Gurumurthy, G. Das, D. Chakraborty, K. Dasgupta, S. Mukherjea, and A. Joshi, "On the structural properties of massive telecom call graphs: findings and implications," in *Proceedings of the 15th ACM international conference on Information and knowledge management*, 2006, pp. 435–444.

[17] A. A. Nanavati, R. Singh, D. Chakraborty, K. Dasgupta, S. Mukherjea, G. Das, S. Gurumurthy, and A. Joshi, "Analyzing the structure and evolution of massive telecom graphs," *IEEE Transactions on Knowledge and Data Engineering*, vol. 20, no. 5, pp. 703–718, 2008.

[18] R. Sarmento, M. Oliveira, M. Cordeiro, S. Tabassum, and J. Gama, "Social network analysis in streaming call graphs," in *Big Data Analysis: New algorithms for a new society*. Springer, 2016, pp. 239–261.

[19] C. Kiss and M. Bichler, "Identification of influencers—measuring influence in customer networks," *Decision Support Systems*, vol. 46, no. 1, pp. 233–253, 2008.

[20] T. Chen, Z. Li, Y. Zhu, J. Chen, X. Luo, J. C.-S. Lui, X. Lin, and X. Zhang, "Understanding ethereum via graph analysis," *ACM Transactions on Internet Technology (TOIT)*, vol. 20, no. 2, pp. 1–32, 2020.

[21] D. Kondor, M. Pósfai, I. Csabai, and G. Vattay, "Do the rich get richer? an empirical analysis of the bitcoin transaction network," *PloS one*, vol. 9, no. 2, p. e86197, 2014.

[22] Y. Wang, L. Wang, and J. Yang, "Egonet based anomaly detection in e-bank transaction networks," in *IOP Conference Series: Materials Science and Engineering*, vol. 715, no. 1. IOP Publishing, 2020, p. 012038.

[23] E. Letizia and F. Lillo, "Corporate payments networks and credit risk rating," *EPJ Data Science*, vol. 8, no. 1, p. 21, 2019.

[24] H. Mao, X. Jin, L. Zhu *et al.*, "Methods of measuring influence of bank customer using social network model," *American Journal of Industrial and Business Management*, vol. 5, no. 04, p. 155, 2015.

[25] R. Kawahara and M. Takeuchi, "Cash flow prediction of a bank deposit using scalable graph analysis and machine learning," in *2021 IEEE International Conference on Big Data (Big Data)*. IEEE, 2021, pp. 1647–1656.

[26] B. Bollobás and B. Béla, *Random graphs*. Cambridge university press, 2001, no. 73.

[27] A.-L. Barabási and R. Albert, "Emergence of scaling in random networks," *science*, vol. 286, no. 5439, pp. 509–512, 1999.

[28] J. Leskovec and A. Krevl, "SNAP Datasets: Stanford large network dataset collection," http://snap.stanford.edu/data, Jun. 2014.

[29] D. J. Watts and S. H. Strogatz, "Collective dynamics of 'small-world'networks," *nature*, vol. 393, no. 6684, pp. 440–442, 1998.

[30] S. N. Dorogovtsev, A. V. Goltsev, and J. F. F. Mendes, "Pseudofractal scale-free web," *Physical review E*, vol. 65, no. 6, p. 066122, 2002.

[31] G. Szabó, M. Alava, and J. Kertész, "Structural transitions in scale-free networks," *Physical Review E*, vol. 67, no. 5, p. 056102, 2003.

[32] R. D. Luce and A. D. Perry, "A method of matrix analysis of group structure," *Psychometrika*, vol. 14, no. 2, pp. 95–116, 1949.

[33] X. Hu, "A data mining approach for retailing bank customer attrition analysis," *Applied Intelligence*, vol. 22, no. 1, pp. 47–60, 2005.

[34] F. Devriendt, J. Berrevoets, and W. Verbeke, "Why you should stop predicting customer churn and start using uplift models," *Information Sciences*, vol. 548, pp. 497–515, 2021.

[35] M. Potamias, F. Bonchi, C. Castillo, and A. Gionis, "Fast shortest path distance estimation in large networks," in *Proceedings of the 18th ACM conference on Information and knowledge management*, 2009, pp. 867–876.

[36] S. Milgram, "The small world problem," *Psychology today*, vol. 2, no. 1, pp. 60–67, 1967.

[37] M. Kamola, E. Niewiadomska-Szynkiewicz, and B. C. Piech, "Reconstruction of a social network graph from incomplete call detail records," in *2011 International Conference on Computational Aspects of Social Networks (CASoN)*. IEEE, 2011, pp. 136–140.