Delft University of Technology

Characterizing temporal bipartite networks - Sequential- Versus cross-tasking

Peters, Lucas J.J.M.; Cai, Juan Juan; Wang, Huijuan

DOI
10.1007/978-3-030-05414-4_3

Publication date
2019

Document Version
Accepted author manuscript

Published in
Complex Networks and Their Applications VII

Citation (APA)
Peters, L. J. J. M., Cai, J. J., & Wang, H. (2019). Characterizing temporal bipartite networks - Sequential- Versus cross-tasking. In L. M. Aiello, C. Cherifi, H. Cherifi, R. Lambiotte, P. Lió, & L. M. Rocha (Eds.), *Complex Networks and Their Applications VII: Proceedings The 7th International Conference on Complex Networks and Their Applications COMPLEX NETWORKS 2018* (1 ed., Vol. 2, pp. 28-39). (Studies in Computational Intelligence; Vol. 813). Springer. https://doi.org/10.1007/978-3-030-05414-4_3

Important note
To cite this publication, please use the final published version (if applicable).
Please check the document version above.

Copyright
Other than for strictly personal use, it is not permitted to download, forward or distribute the text or part of it, without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license such as Creative Commons.

Takedown policy
Please contact us and provide details if you believe this document breaches copyrights.
We will remove access to the work immediately and investigate your claim.

This work is downloaded from Delft University of Technology.
For technical reasons the number of authors shown on this cover page is limited to a maximum of 10.

Characterizing Temporal Bipartite Networks - sequential- vs cross-tasking

Lucas J.J.M. Peters[1], Juan-Juan Cai[2], and Huijuan Wang[2]

[1] Faculty of Applied Sciences, Delft University of Technology, The Netherlands,
[2] Faculty of Electrical Engineering, Mathematics, and Computer Science, Delft University of Technology, The Netherlands,
h.wang@tudelft.nl

Abstract. Temporal bipartite networks that describe how users interact with tasks or items over time have recently become available. Such temporal information allows us to explore user behavior in-depth. We propose two metrics, the relative switch frequency and distraction in time to measure a user's sequential-tasking level, i.e. to what extent a user interacts with a task consecutively without interacting with other tasks in between. We analyze the sequential-tasking level of users in two real-world networks, an user-project and an user-artist network that record users' contribution to software projects and users' playing of musics from diverse artists respectively. We find that users in the user-project network tend to be more sequential-tasking than those in the user-artist network, suggesting a major difference in user behavior when subject to work related and hobby-related tasks. Moreover, we investigate the relation (rank correlation) between the two sequential-tasking measures and another 10 nodal features. Users that interact less frequently or more regularly in time (low deviation in the time interval between two interactions) or with fewer items tend to be more sequential-tasking in the user-project network. No strong correlation has been found in the user-artist network, which limits our ability to identify sequential-tasking users from other user features.

1 Introduction

Bipartite networks [2] are networks whose nodes can be divided into two disjoint sets $\mathbb{U}$ and $\mathbb{I}$ such that every link connects a node in $\mathbb{U}$ and a node in $\mathbb{I}$. Bipartite networks have been widely used to represent, for example, which author has contributed to which paper, which user has listened to which music and the ownership between banks and assets. Using the information that is nowadays continuously created, shared and recorded in the social networks, we could obtain the temporal features of the interactions between the two disjoint sets of a bipartite network. Temporal bipartite networks record not only e.g. which user listened to which music but also at which time stamp(s). A link in a static bipartite network is enriched in a temporal bipartite network as a set of contacts/interactions (temporal links) over time between two nodes from the two sets (user set and item set) respectively. Such temporal information is essential to explore user behavior. The items in the item set, for example projects and artists, can be regarded as tasks, in this case work or hobby related. In this paper, we exploit the temporal information to distinguish a user's behavior between sequential-tasking and cross-tasking. A

cross-tasking user tends to interact with other tasks/items in between interactions with a given task while a sequential-tasking user contributes consecutively to a task without the interruption of any other task.

With the temporal information of each interaction, also called contact, we aim to address: How to quantify a user's sequential or cross-tasking level? Which kind of users are more sequential-tasking? To this end, we propose two metrics to characterize the sequential-tasking level of a user and explore how the sequential-tasking level is related to other temporal features of a user. We illustrate our methods in two real-world temporal bipartite networks: an user-artist network that records the timestamps when a user plays a song of an artist and an user-project network that describes the moments when a software developer contributes to a software project. We choose these two networks, also motivated by the question: Do human behave differently in listening to music for pleasure and in contributing to working tasks like software development?

Methods to characterize features of a temporal network have been recently studied. [3, 4, 8] How temporal network features affect a diffusion process has also been explored. It has been shown that such effect can not be captured by the static network, which integrates a temporal network over time[5–7]. In view of the existing literature, we highlight that the contribution of this paper is threefold. First, we contribute new methodologies to characterize nodal level temporal features with respect to how a node switch tasks or interactions with items over time, beyond the classic centrality metrics that describes static node features [9]. Second, we unravel as well the relationships among diverse temporal centrality metrics. Third, bipartite networks are a special type of multi-layer network [10–13]. Our work may inspire the characterization of multi-layer temporal networks in general. The paper is organized as follows. In Section 2, we introduce the two real-world datasets, their bipartite temporal network representation and basic statistical properties. We propose two measures to quantify a user's sequential-tasking level and compare the two features in the two networks in Section 3. The relation between the proposed nodal features are compared with another ten nodal properties in Section 4 to explore which kind of nodes tend to be sequential-tasking. Conclusions and future challenges are addressed in Section 5.

2 Real-world Temporal Bipartite Networks

In this section, we introduce the construction of the two real-world temporal bipartite networks that will be studied as examples throughout this work and their basic statistical features.

2.1 Dataset description

The user-artist dataset records the music playing activities of 1000 random users in last.fm [1]. A temporal contact (u,i,t) between a user u and an artist i at time t represents that user u plays a song of artist i at time t. The user-project dataset contains the timestamps when an user contributes to a project (resolves a thread in the Linux kernel)[15]. Since we aim to explore the sequential-tasking level of a user, i.e. the extent to which a user interacts consecutively with an item without interacting with other items, we consider only the users that interact two or more times with at least one item. In other words, we remove users that interact

maximally once with any item, for whom it is impossible to distinguish between sequential- and cross-tasking. The resultant number of users, the number of items and the observation time window for each dataset are given in Table 1.

2.2 Network representation

A temporal bipartite network measured within time window $[1,2,...,T]$ and composed of a set $\mathbb{U}$ of U users and a set $\mathbb{I}$ of I items can be represented by a $U \times I \times T$ temporal adjacency matrix $\mathscr{A}$. An element $\mathscr{A}(u,i,t) = 1$ or $\mathscr{A}(u,i,t) = 0$ indicates, respectively, that there is a contact or no contact between the user u and item i at time step t, where $u = 1,\ldots,U$, $i = 1,\ldots,I$ and $t = 1,\ldots,T$. Two aggregated networks can be derived from a temporal network: The weighted aggregate network can be represented by a weighted adjacency matrix $\mathbf{W}$ where the element W_{ui} equals the number of interactions between user $u \in \mathbb{U}$ and item $i \in \mathbb{I}$; The unweighted aggregated network can be represented by the unweighted adjacency matrix $\mathbf{A}$, where $A_{ui} = 1$ if user u and item i interact at least once and $A_{ui} = 0$ if u and i do not interact. Both matrix W and A are of size $U \times I$.
The total number of interactions of a user u is denoted by s_u and $s_u = \sum_{i=1}^{I} W_{ui}$ is actually the node strength in the weighted network W [14]. The number of distinct items a user u interacts with is denoted by d_u and $d_u = \sum_{i=1}^{I} A_{ui}$ is the nodal degree in the unweighted network A. Likewise, the total number of interactions of an item i is given by $s_i = \sum_{u=1}^{U} W_{ui}$ and the the number of distinct users that are linked to an item i is $d_i = \sum_{u=1}^{U} A_{ui}$.

2.3 Basic network characteristics

	user-project network	user-artist network
#users U	$13,990$	986
#items I	$330,051$	$83,982$
#temporal links/contacts	$979,846$	$16,964,897$
observation time window	Mar. 1993 - May 2014	Feb. 2005 - Sep. 2013
avg #interactions/user	70	$17,206$
avg #items/user	41	748

Table 1: Basic properties of the user-project and user-artist networks.

We aim to design metrics to measure the sequential- or cross-tasking level of a user, which allow us to compare users. We explore first how much users differ from each other in e.g. the number of interactions, the number of distinct items a user interacts with. Such difference or similarity among users are crucial for the metric design with respect to which kind of normalizations should be taken into account such that users can be well compared. As shown in Figure 1 and 2, the distributions of the number of interactions s_u of a user and the distinct items d_u that a user interacts with are highly heterogeneous in both networks. Such high heterogeneity can be also observed among the items in the number of interactions s_i per item and the number of distinct users d_i an item interacts with and among the user item pair in the number of interactions w_{ui} per user and item pair. These

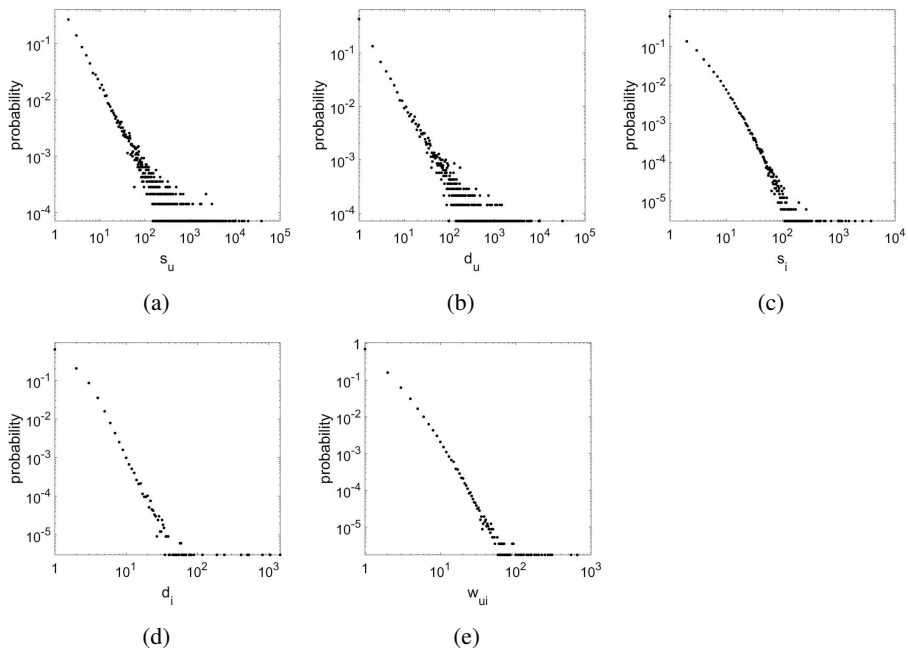


Fig. 1: The distributions of **(a)** the number of interactions s_u per user, **(b)** the number of distinct items d_u that a user interacts with, **(c)** the number of interactions s_i per item, **(d)** the number of distinct users d_i an item interacts with and **(e)** the number of interactions w_{ui} per user and item pair when $w_{ui} > 0$ in the user-project network.

observations imply that normalizations are essential in the design of the metrics to compare users.

3 Quantifying the sequential/cross-tasking level

For an arbitrary user, his interactions or contacts with items can be represented by a vector C of length n where element $C(j)$ is the index of the j-th contacted item in time and n is the total number of contacts of the user. Examples of the contact vectors are given in Figure 3. We measure the sequential-tasking level of a user based on the sequence C. The real time delay between any two consecutive contacts e.g. the j-th and the $(j+1)$-th contacts is ignored (or regarded as 1 time step) because during the period between two consecutive contacts, a user is expected to participate in other activities that are not covered in our datasets.

We propose two metrics to measure the sequential-tasking level from two perspectives: I. the relative switch frequency, i.e. the scaled number of times that a user switches from a task/item to another over time and II. the relative distraction i.e. the fraction of time stamps that is used to contact/contribute to other tasks between the first and the last time stamp of contact with a given task.

From the given contact sequence C we could deduce the number m of distinct items the user interacts with and the number n_i of times the user interact with

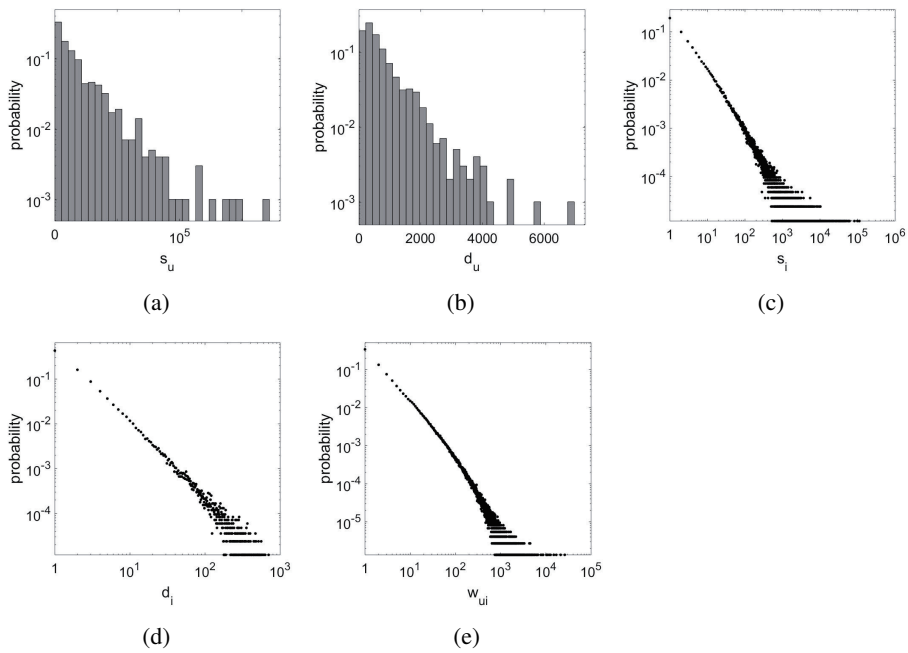


Fig. 2: The distributions of **(a)** the number of interactions s_u per user, **(b)** the number of distinct items d_u that a user interacts with, **(c)** the number of interactions s_i per item, **(d)** the number of distinct users d_i an item interacts with and **(e)** the number of interactions w_{ui} per user and item pair when $w_{ui} > 0$ in the user-artist network.

each item i, $1 \leq i \leq m$. In this case, the user has in total $n = \sum_{i=1}^{m} n_i$ interactions. The item contact frequency vector $\{n_1, n_2, , , , n_m\}$ tells the number of contacts that each of the m items receives.

3.1 Relative switch frequency θ

We first quantify the sequential-tasking level from the perspective of the relative switch frequency, i.e. the scaled number of times that a user switches from a task/item to another over time.

Given a contact sequence C, the total number of switches $\Theta = \sum_{i=1}^{n-1} 1_{C(i) \neq C(i+1)}$ can be computed by examining every two consecutive contacts $C(i)$ and $C(i+1)$ and counting how many times $C(i)$ and $C(i+1)$ are different i.e. when the indicator function $1_{C(i) \neq C(i+1)}$ is 1. Examples are given in Figure 3.

Users differ evidently in the number of distinct items they contact and the number of total interactions thus differ also in the interacting frequency vector $\{n_1, n_2, , , , n_m\}$. Trivially, the interacting frequency vector influences the possible range and distribution of the number of switches when considering all possible contact sequences that follow the same interacting frequency vector. Given the number m of distinct items a user contacts, the minimal number of switches is $\Theta = m - 1$, which corresponds to the highest possible sequential-tasking level. The sequence C_1 in Figure

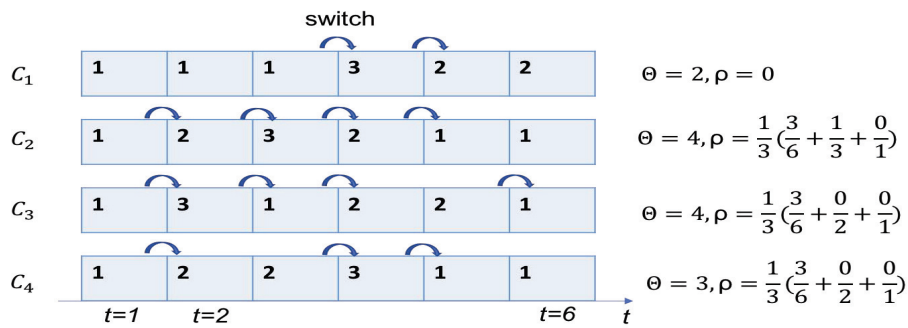


Fig. 3: Four interaction/contact sequence examples of a user that interacts with $m = 3$ items $\{1,2,3\}$, $n_1 = 3$, $n_2 = 2$ and $n_3 = 1$ times respectively. The number of switches Θ and the distraction level ρ are given for each example.

3 is one such example. Given the number of items m a user contacts, more interactions tends to lead to more switches. Hence, we propose to consider the following normalized number of switches

$$\theta = \frac{\Theta - \Theta_{min}}{\Theta_{ran} - \Theta_{min}} \tag{1}$$

where $\Theta_{min} = m - 1$ is the minimal possible number of switches and Θ_{ran} is the average number of switches when the contact sequence (the time order of the contacts) is randomly shuffled. The normalized number of switches of a contact sequence counts the number of switches that exceeds the minimal possible number of switches, relative to the case when the contacts are randomly ordered in time but following the same item contact frequency vector. A smaller normalized number θ of switches suggests a more sequential interaction with the items/tasks. Given the item contact frequency vector $\{n_1, n_2, , , , n_m\}$, the average number Θ_{ran} of switches when the time order of the contact sequence is randomly shuffled can be approximated by

$$\Theta_{ran} \sim (n-1) \sum_{i=1}^{m} \frac{n_i}{n} (1 - \frac{n_i - 1}{n}) \tag{2}$$

Next we give the reasoning for this approximation. Let $B_1 = 1_{\text{[there is a switch after the 1st contact]}}$. Clearly, B_1 is a Bernoulli random variable with success probability p given by:

$$p = Pr[B_1 = 1] = \sum_{i=1}^{m} Pr[C(2) \neq i | C(1) = i] Pr[C(1) = i] = \sum_{i=1}^{m} \frac{n_i}{n} (1 - \frac{n_i - 1}{n})$$

Similarly, we define B_j, $j = 2, \ldots, n-1$ indicating the switches at all possible contacts. Assuming that $B_2, \ldots, B_{n-1}$ are independently and identically distributed as B_1, that is from Bern(p), we obtain the expected number of total switch: $\sum_{i=1}^{n-1} E(B_j) = (n-1)p$, which is exactly the right hand side of (2). This quantity is an approximation of Θ_{ran} because the assumption that B_j's being independently and identically distributed is generally not satisfied. The approximation

is precise if every element n_i in the item contact frequency vector is infinite, so is the total number of contacts.
Further, we evaluate the precision of this approximation by comparing it with the Θ^*_{ran} obtained from our simulations. We choose from each dataset a set of users with diverse activities profiles (in the number of interactions n and number of distinct items m it interacts with) and consider the item contact frequency vectors of these users to verify our approximation. We consider users that interact to at least two items. We first classify the numbers of interactions of users as low, medium or high such that equal amount of users can be classified to each of the three categories. Independently and similarly, we classify the the number of distinct items a user interacts with as low, medium or high. We obtain 9 categories when combining these two classifications. In the user-project dataset, we randomly choose one user per category. In the user-artist dataset one category has no users and we randomly choose two users from each of the 8 non-empty categories. For each selected user, we derive the item contact frequency vector and compute the average number of switches Θ^*_{ran} of the h realizations of the randomly shuffled contact sequences. The average ratio of Θ_{ran} obtained by (2) over Θ^*_{ran} is obtained for each dataset. As shown in Table 2, the precision $\frac{\Theta^*_{ran}}{\Theta_{ran}}$ is relatively high in both networks and stable as the number of realizations h increases (quadruples). The precision is higher in user-artist network. As explained earlier, our approximation tends to be more precise if the number of interactions per user-item pair is larger. The significantly larger number of interactions per user in the user-artist network than that in the user-project network as shown in Table 1 supports why our approximation in user-artist network is more precise.

	h=125	h=500
user-project network	0.8996	0.9040
user-artist network	0.99985	0.99991

Table 2: The average ratio of the Θ_{ran} obtained by Eq. (2) to that Θ^*_{ran} obtained by simulations over all the selected users in each dataset. Per user, Θ^*_{ran} is the average over h realizations of the randomized contact sequences.

3.2 Relative distraction ρ in time

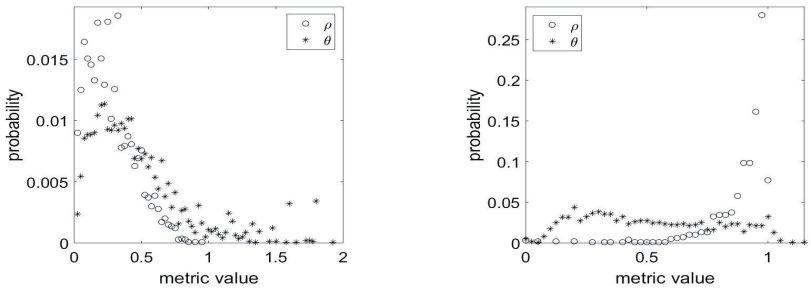
To what extent a user sequentially interacts with items or processes tasks can also be measured according to the fraction of time stamps that a user is distracted by other tasks in finishing a task, between the first and the last interaction with that given task. We use $p_f(i) = \min_{t,1\leq t\leq n}\{C(t) = i\}$ to denote the time stamp when the user under consideration interacts with item i for the first time. Similarly, $p_l(i) = \max_{t,1\leq t\leq n}\{C(t) = i\}$ denotes the time index when the user interacts with item i for the last time. We measure the relative distraction as

$$\rho = \frac{1}{m}\sum_{i=1}^{m}\frac{\sum_{k=p_f(i)}^{k=p_l(i)} 1_{\{C(k)!=i\}}}{n_i + \sum_{k=p_f(i)}^{k=p_l(i)} 1_{\{C(k)!=i\}}} \tag{3}$$

where $\sum_{k=p_f(i)}^{k=p_l(i)} 1_{\{C(k)!=i\}}$ counts the number of other items occur in between the first and last interaction with item i and the ratio implies the fraction of time

disrupted by other items in completing item i. The total time to complete item i is $n_i + \sum_{k=p_l(i)}^{k=p_l(i)} 1_{\{C(k)!=i\}}$, which equals $p_l(i) - p_f(i) + 1$.
The relative distraction level ranges within $0 \leq \rho < 1$ and a smaller ρ suggests a high sequential-tasking level, i.e. consecutively interacting with one task without the interactions with any other task in between (see e.g. Figure 3)

3.3 Comparison of two real-world networks



(a) (b)

Fig. 4: Probability distribution of the relative switch frequency θ and relative distraction in time ρ in the **(a)** user-project and **(b)** user-artist network

We compute the relative switch frequency θ and the relative distraction in time ρ for each user in these two networks. The probability distribution for each measure is plotted for each network in Figure 4. Both measures follow a similar distribution in the user-project network, where many users are sequential-tasking and few are cross-tasking. In the user-artist network, the relative switch frequency tends to be homogeneously distributed whereas many users are highly distracted in time. According to the definition of the distraction ρ level, a user that interacts many times with many items may tend to be more distracted. The distributions of the number of interactions and items per user are heterogeneous in the user-artist network, as shown in Figure 2: many (few) users have a small (large) number of interactions/items. Hence, the large number of users with a high distraction level reveals intrinsic feature of the user-artist network, beyond the relative switch frequency and the number of interactions/items can describe. Users tend to be sequential-tasking for tasks like software programming but cross-tasking in hobby type tasks like listening to musics.

4 Correlation between sequential/cross-tasking level and other centrality metrics

We address further the question which kind of users in a network tend to be sequential- or cross-tasking. We investigate thus the correlations between the measures we proposed and a large set of basic centrality metrics that describe

various temporal nodal features. We propose to consider the following centrality metrics for a user u in a generic temporal bipartite network:

- the number of interactions s_u
- the number of distinct items d_u a user interacts with
- the average weight (number of interactions) with an item that the user u interacts with $E[W_u] = \frac{\sum_i W_{ui} A_{ui}}{d_u}$
- the normalized standard deviation of the weight (number of interactions) with an item the user u interacts with $\frac{\sigma[W_u]}{E[W_u]} = \frac{\sqrt{\frac{\sum_i (W_{ui} - E[W_u])^2 A_{ui}}{d_u}}}{E[W_u]}$
- the average inter-arrival time $E[\tau]$ of a user, where the inter-arrival time is the time delay between two consecutive interactions of a user
- the normalized standard deviation $\sigma[\tau]/E[\tau]$of the inter-arrival time of a user
- the average number of users of the items that a user u interacts with $\frac{\sum_i d_i A_{ui}}{d_u}$: compute for each item that the user u interacts with, the number of users this item interacts with and compute the average number of users over all the items that u interacts with
- the average number of interactions of the items that a user interacts with $\frac{\sum_i s_i A_{ui}}{d_u}$
- the normalized standard deviation of the number of users of the items that user u interacts with $\sqrt{\frac{\sum_i (d_i - \frac{\sum_i d_i A_{ui}}{d_u})^2 A_{ui}}{d_u}} / \frac{\sum_i d_i A_{ui}}{d_u}$
- the normalized standard deviation of the number of interactions of the items that user u interacts with $\sqrt{\frac{\sum_i (s_i - \frac{\sum_i s_i A_{ui}}{d_u})^2 A_{ui}}{d_u}} / \frac{\sum_i s_i A_{ui}}{d_u}$

The last four centrality metrics explore the neighbors that are two hops away from the user u. The average and normalized standard deviation of the inter-arrival time explore new temporal features of a user beyond the two proposed metrics.

We compute the above ten centrality metrics as well as the two measures we proposed earlier for each node in each real-world network. Within each network, we rank the nodes according to each of the 12 metrics and compute the rank correlation between any two metrics, which are given in Table 3. The relative switch frequency θ and distraction ρ are strongly and positively correlated in the user-project network whereas the correction is less strong in the user-artist network, qualitatively explaining part of the behavior in Figure 4. In the user-project network, less active users with respect to the number of interactions or items and users that interact more regularly (low standard deviation in inter-arrival time) tend to be more sequential-tasking. However, the sequential-tasking level θ and ρ are not strongly correlated with the other user features in the user-artist network. Interestingly, the last four metrics in the list and Table 3, i.e. the average and normalized standard deviation of the number of users/interactions of the items a user interacts with are strongly correlated in the user-artist network. If a user interacts with items that are popular regarding to the number of users/interactions they attract on average, these items tend to be similar in the number of users/interactions they attracted. However, the contrary has been observed in the user-project network: if a user interact with popular items on average, these items are heterogeneous in their popularity.

In the user-project network, users are on average more sequential-tasking and sequential-tasking users tend to be those less active and regular in time in interactions. For hobby type tasks, users tend to be cross-tasking and the sequential-tasking level of a users cannot be distinguished from the 10 user features.

5 Conclusion

The two metrics that we proposed, the relative switch frequency and distraction in time, quantify the sequential-tasking level of a user from two perspectives. Using the user-project and user-artist networks as examples, we illustrate the evidently different sequential-tasking levels when users are subject to work-related and hobby-related tasks. Users tend to be more sequential-tasking in user-project network. Our correlation study between the two sequential-tasking level measures and another 10 nodal centrality metrics unravel the possibility to identify sequential- or cross-tasking users from other features in the user-project network but not the user-artist network. Our work is deemed as the starting point to investigate the new type of centrality metrics that exploit a user's temporal activities or time series. It is interesting to explore the relation between such centrality metrics as well as those information-theoretic measures proposed in the study of human behavior patterns [16, 17].

References

1. www.last.fm
2. Tao Zhou, Jie Ren, Mat Medo, and Yi-Cheng Zhang: Bipartite network projection and personal recommendation, Phys. Rev. E 76, 046115 (2007)
3. Holme P and Saramäki:Temporal networks, *Physics reports* **519** 97–126 (2012)
4. Holme P: Modern temporal network theory: a colloquium, *The European Physical Journal B* **88** 234 (2015)
5. Karsai M, Kivelä M, Pan R K, Kaski K, Kertész J, Barabási A L and Saramäki: Small but slow world: How network topology and burstiness slow down spreading, *Physical Review E* **83** 025102 (2011)
6. Lambiotte R, Tabourier L and Delvenne J C: Burstiness and spreading on temporal networks *arXiv preprint arXiv:1305.0543* (2013)
7. Moinet A, Starnini M and Pastor-Satorras R: Burstiness and Aging in Social Temporal Networks, *Physical review letters* **114** 108701 (2015)
8. Xiu-Xiu Zhan, Alan Hanjalic, and Huijuan Wang: Information diffusion backbones in temporal networks, arXiv preprint arXiv:1804.09483 (2018)
9. Cong Li, Qian Li, Piet Van Mieghem, H. Eugene Stanley and Huijuan Wang: Correlation between centrality metrics and their application to the opinion model, The European Physical Journal B **88** 65 (2015)
10. Huijuan Wang, Qian Li, Gregorio D, Agostino, Shlomo Havlin, H Eugene Stanley and Piet Van Mieghem: Effect of the interconnected network structure on the epidemic threshold, Physical Review E **88** 022801 (2013)
11. Xuqing Huang, Shuai Shao, Huijuan Wang, Sergey V Buldyrev, H Eugene Stanley and Shlomo Havlin, The robustness of interdependent clustered networks, Europhysics Letters**101** 18002 (2013)
12. Meng Liu, Daqing Li, Pengju Qin, Chaoran Liu, Huijuan Wang and Feilong Wang,Epidemics in interconnected small-world networks, PloS one **3**e0120701 (2015)
13. De Domenico, Manlio and Solé-Ribalta, Albert and Cozzo, Emanuele and Kivelä, Mikko and Moreno, Yamir and Porter, Mason A. and Gómez, Sergio and Arenas, Alex, Mathematical Formulation of Multilayer Networks,Phys. Rev. X**3**041022 (2013)

	ρ	θ	num. interac.	num. item	avg weight	std weight	avg inter-arriv. time	std inter-arriv. time	avg num. users of item	avg num. int. of item	std num. users of item	std num. int. of item
user-project dataset												
ρ	**100**	**99.7**	**67**	**69**	-12	28	15	**59**	-20	-20	22	34
θ		**100**	**66**	**68**	-12	25	15	**56**	-19	-20	19	29
num. interac.			**100**	**88**	9	**52**	41	**79**	-23	-21	41	**58**
num. items				**100**	-31	35	**52**	**74**	-25	-29	42	**58**
Avg weight					**100**	56	-19	-13	13	25	1	1
std weight						**100**	-18	45	9	18	30	44
avg inter-arriv. time							**100**	15	9	6	5	-9
std inter-arriv. time								**100**	-20	-19	30	44
avg num. users of items									**100**	**96**	42	20
avg num. int. of items										**100**	45	28
std num. users of items											**100**	**86**
std num. int. of items												**100**
user-artist dataset												
ρ	**100**	**67**	38	47	10	25	-29	12	-11	-19	4	16
θ		**100**	-11	18	-35	-17	14	1	-2	-9	-5	6
num. interac.			**100**	**65**	**72**	**55**	**-76**	8	-34	-34	31	33
num. items				**100**	2	37	-47	0	**-52**	**-59**	44	**57**
avg weight					**100**	44	**-58**	14	1	8	4	-5
std weight						**100**	-42	8	-14	-15	15	18
avg inter-arriv. time							**100**	19	30	29	-30	-30
std inter-arriv. time								**100**	10	10	-11	-10
avg num. users of items									**100**	**97**	**-95**	**-92**
avg num. int. of items										**100**	**-88**	**-91**
std num. users of items											**100**	**93**
std num. int. of items												**100**

Table 3: Rank correlation between any two of the 12 user features in the user-project and user-artist networks in percentages.

14. Huijuan Wang, Linda Douw, Javier Hernández and Jaap Reijneveld, Kees Stam and Piet Van Mieghem, Effect of tumor resection on the characteristics of functional brain networks, Phys. Rev. E **82**, 021924 (2010)
15. Linux kernel mailing list threads network dataset – KONECT (2017)
16. Song, Chaoming and Qu, Zehui and Blumm, Nicholas and Barabási, Albert-László: Limits of Predictability in Human Mobility, Science **327**, Issue 5968, 1018–1021 (2010)
17. Takaguchi, Taro and Nakamura, Mitsuhiro and Sato, Nobuo and Yano, Kazuo and Masuda, Naoki, Predictability of Conversation Partners, Phys. Rev. X **1** 011008 (2011)