Applied Network Science

RESEARCH

Open Access

From co-location patterns to an informal social network of gig economy workers

Gustavo Pilatti[1,2*], Cristian Candia[3,4,5,6], Alessandra Montini[1] and Flávio L. Pinheiro[2]

*Correspondence: pilatti@usp.br

[1] Faculdade de Economia, Administração e Contabilidade, Universidade de São Paulo, Av. Prof. Luciano Gualberto, 908, São Paulo 05508-010, Brazil
[2] NOVA Information Management School (NOVA IMS), Universidade Nova de Lisboa, Campus de Campolide, 1070-312 Lisbon, Portugal
[3] Computational Research in Social Science Laboratory, School of Engineering and School of Government, Universidad del Desarrollo, Santiago 7610658, Chile
[4] Centro de Investigación en Complejidad Social, Facultad de Gobierno, Universidad del Desarrollo, Santiago 7610658, Chile
[5] Instituto de Data Science, Facultad de Ingeniería, Universidad del Desarrollo, 7610658 Las Condes, Chile
[6] Northwestern Institute on Complex Systems, Kellogg School of Management, Northwestern University, Evanston, IL 60208, USA

Abstract

The labor market has transformed with the advent of the gig economy, characterized by short-term and flexible work arrangements facilitated by online platforms. As this trend becomes increasingly prevalent, it presents unique opportunities and challenges. In this manuscript, we comprehensively characterize the social networks of gig economy workers in each of the 15 cities studied. Our analysis reveals a scaling relationship between networks and the city population. In particular, we note the high level of modularity of the networks, and we argue that it results from the natural specialization of couriers along different areas of the cities. Furthermore, we show that degree and betweenness centrality is positively correlated with income but not with tenure. Our findings shed new light on the social organization of the gig economy workers and provide valuable insights for the management and design of gig economy platforms.

Keywords: Gig economy, GPS data, Big data, Co-location social network, Complex networks

Introduction

The transition of the global economy to a service-driven model has resulted in services playing a crucial role in economic growth, as well as corporate and personal well-being and consumption patterns (Evans and Gawer 2016; Vallas and Schor 2020). This shift has been further accelerated by technological advancements and innovative business models that enable consumers to access service-related transactions without requiring ownership or long-term commitment (Weidenstedt et al. 2023). As a result, labor markets are witnessing the rise of short-term and flexible work arrangements mediated by online platforms (Woodcock and Graham 2019). Specifically, food-delivery couriers in the "gig economy" perform tasks that are often geographically dispersed and require mobility and coordination. However, the social networks of these workers are primarily hidden and inaccessible, as they rely on offline interactions and lack traceable digital footprints. Moreover, these workers have significant idle time while logged in to the platform, waiting for new deliveries demand or resting between deliveries. During this idle time, offline interactions and information exchange with other couriers are highly likely to occur, forming and strengthening their social ties. By understanding the social networks of gig economy workers, valuable insights can be obtained for both workers and platforms, such as how they spread innovations, coordinate behaviors, and their well-being (Perren and Kozinets 2018). This study proposes to use anonymized data on couriers' co-location patterns to investigate their informal social network.

Gig economy workers are often invisible to traditional labor data sources, as their work is mediated by online platforms that may obscure their activities and interactions (Cherry 2016; Weidenstedt et al. 2023). While analyzing labor patterns requires clear and objective data so that statistical measurements can be effectively performed, traditional approaches will be of little benefit if the elements contributing to professional success cannot be correctly quantified. The literature on gig economy workers is primarily theoretical or based on small sample surveys (Md Fadzil and Che Azmi 2022; Yeganeh 2021; Weidenstedt et al. 2023; Wood et al. 2019). Therefore, network theory can provide a more comprehensive understanding of the gig economy labor market by analyzing the complex interactions and relationships between them. Since the platforms-as-market are important economic organizations inside the service-driven economy and social networks may co-exist with this business type, analyzing the network structure inside these organizations makes it possible to understand how it shapes the gig economy workforce and impacts its success. Nevertheless, the literature affirms that social structures may be inferred from user interactions on a network rather than from explicitly declared relationships (Tubaro 2021). Still, it will lead to a low signal-to-noise ratio (Gupte and Eliassi-Rad 2012). One alternative is to look for proxies that can help us infer the existence of such relationships. In that sense, the co-location of individuals in space and time has been a popular approach to inferring social networks (Njoo et al. 2018; Hsieh et al. 2015; Lu et al. 2022; Njoo et al. 2017; Pi et al. 2018). Such an approach enables the capture of relationships at large between individuals without relying on subjective surveys or self-reports (Lu et al. 2022). Previous works have shown that the co-location of individuals (i.e., their simultaneous presence in the same location) can indicate the existence of a social relationship, as friends tend to visit similar places and meet face-to-face (Crandall et al. 2010; Psorakis et al. 2012; Hsieh et al. 2015). Hence, allows studying related phenomena such as the spread of social influence and contagion, which also take spatial proximity as the conduit of social influence, competition, normative legitimization (experience and enforce norms), and social learning (Iyengar et al. 2011). The core challenge in inferring a network from co-location patterns is to identify appropriate controls to distinguish between coincidental and recurrent ties when dealing with measures such as frequency, correlation, similarity, probability, or importance (Njoo et al. 2018).

Here, we use methods from network analysis to study informal social networks inferred from the co-location patterns of couriers from a popular food delivery platform in Brazil. In that sense, we provide a new perspective to understand the gig economy and the underlying social organization of gig workers, expanding on past works on the same domain (Kinder et al. 2019; Huang et al. 2019) and contributing to deepening the knowledge about this new economy. We compare the networks obtained from 15 different cities in Brazil, characterize their features, and show that the emergent features of these networks seemingly scale with population size. Finally, we show that the courier's location in such networks can constitute a valuable predictor of couriers' performance when measured by income (as a proxy for performance) (Wood et al. 2019). In particular, couriers with a higher degree and betweenness centrality can obtain more income than the ones with the lowest centrality.

Data and methods

Data origin

We utilized data obtained from a prominent Brazilian online delivery mobile phone application in which registered users can order food and groceries from listed merchants. Couriers, who have profiles on the platform, carry out the deliveries. Our study maps the informal social network of gig workers who operate through the platform (Pilatti et al. 2023) following the process described in Graphical Abstract—Fig. 1.

To achieve this goal, we collected telemetry data pertaining to the geo-locations of couriers who worked in 15 cities across Brazil for 14 days in February 2022. This information is our original dataset and will be described below. The selected cities vary in population size and are located in different regions of Brazil. More information about cities can be found in Table 1. Furthermore, the chosen period was selected while considering conditions such as weather, epidemic outbreaks, and political situations. Information about the couriers and their locations was anonymized to maintain confidentiality. It is important to note that gig workers use the platform solely for work-related activities, and any interactions or activities outside the app are not available for analysis. For instance, conversations with other couriers on messaging apps and data from their mobile phones are not recorded or stored and therefore are not accessible for study.

Data treatment

Our original dataset ("raw data") is composed of geo-located observations, each representing the geographic coordinates (latitude and longitude) of a courier location at each time (day, hour, minute, and second). These coordinates, rounded to the fifth decimal point, afford an accuracy of approximately one meter. The location data was reported on an average frequency of every fifteen seconds and collected via the couriers' mobile application. This expansive data collection yielded a total of 65,405,061 recorded events.

For our analysis, we aggregated the observations into discrete, ten-minute temporal windows. This methodological choice was driven by our objective of accurately estimating informal social ties that may be inferred from courier co-location patterns. Consequently, we implemented a series of measures designed to isolate events that likely corresponded to periods of courier inactivity, such as rest periods, meal breaks, or instances of waiting for new delivery requests.

To minimize the potential for false positives arising from transient co-location, we imposed a two-minute stationary threshold within the ten-minute window. This

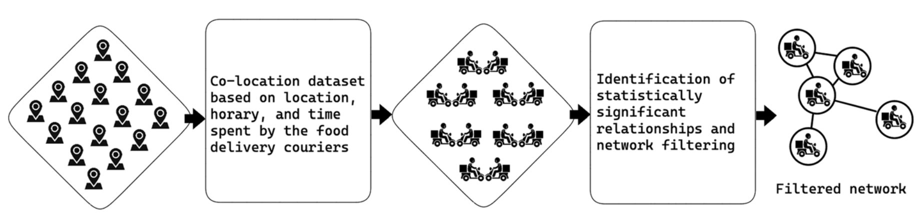

Fig. 1 Proposed framework to study couriers co-location and tie significance

exclusion criterion discarded instances where couriers did not remain in a single location for at least two minutes, effectively filtering out potential anomalies. Examples of such anomalies include scenarios where couriers momentarily shared a location due to traffic constraints, simultaneous stops at traffic signals, or collection/delivery of orders at identical addresses. This precautionary step enhanced the reliability of our findings, improving the logical rigor of our analytic process.

Moreover, to avoid noise from new couriers starting or leaving their activities through the application during the 14 days of analysis, only couriers that remained operative during the 14 days before or after the studied period were considered. By being operative, we considered at least one order delivered in the period. This way, the final dataset that was used in the study is composed only of operative couriers that have met other couriers during the time of the study, and it is composed of a unique courier identification code, the time slot, and the geographic coordinate where this courier met others couriers.

Finally, courier locations were matched to obtain information on couriers' temporal (day, hour, and minutes rounded in 10 min time slot) and spatial (geographic coordinates up to the 5 decimal cases to get a precision of 1-meter radius) co-location. Note that at this stage, the only events we have concerned couriers that were stopped in a location at a given time. Figure 2 shows the geographical co-location of the couriers, while Fig. 3A is an illustrative example that summarizes how the data was cleaned and which information was kept in the final dataset. This dataset informs of 100,294 unique co-location events of 19,207 unique couriers, segmented across the 15 cities under study.

Courier–courier informal network

For each city under analysis, we construct an undirected and weighted network. Each node represents a courier, and each link indicates the number of co-location events observed between a pair of couriers in time and space. A link with a zero weight indicates that such a pair of couriers were never observed to co-locate. The network was built on top of the final dataset presented earlier, using as a matching key the time slot and the geographic coordinate to combine the couriers in pairs or groups (i.e., if two rows of the final dataset had the same time slot and the same geographic coordinate, it means that both couriers were at the same place, at the same time, so, an edge between the two were build).

To estimate which ties exhibit a positive and statistically significant co-location pattern, we use the ϕ-correlation method to estimate the strength of the relationship between couriers. We then control for events attributed to pure chance (Ronen et al. 2014; Candia et al. 2019; Kalgotra et al. 2020; Candia et al. 2022). We define ϕ_{ij} as the ϕ -correlation coefficient between a pair of couriers i and j, which can be computed as:

$$\phi_{ij} = \frac{a_{ij}Z - a_i a_j}{\sqrt{a_i a_j (Z - a_i)(Z - a_j)}}, \tag{1}$$

where the ϕ-correlation coefficient ϕ_{ij} is computed using the number of observed co-locations between i and j (a_{ij}), the number of events in which i appears (a_i), and the total number of events observed (Z). Positive and negative values of ϕ_{ij} indicate that

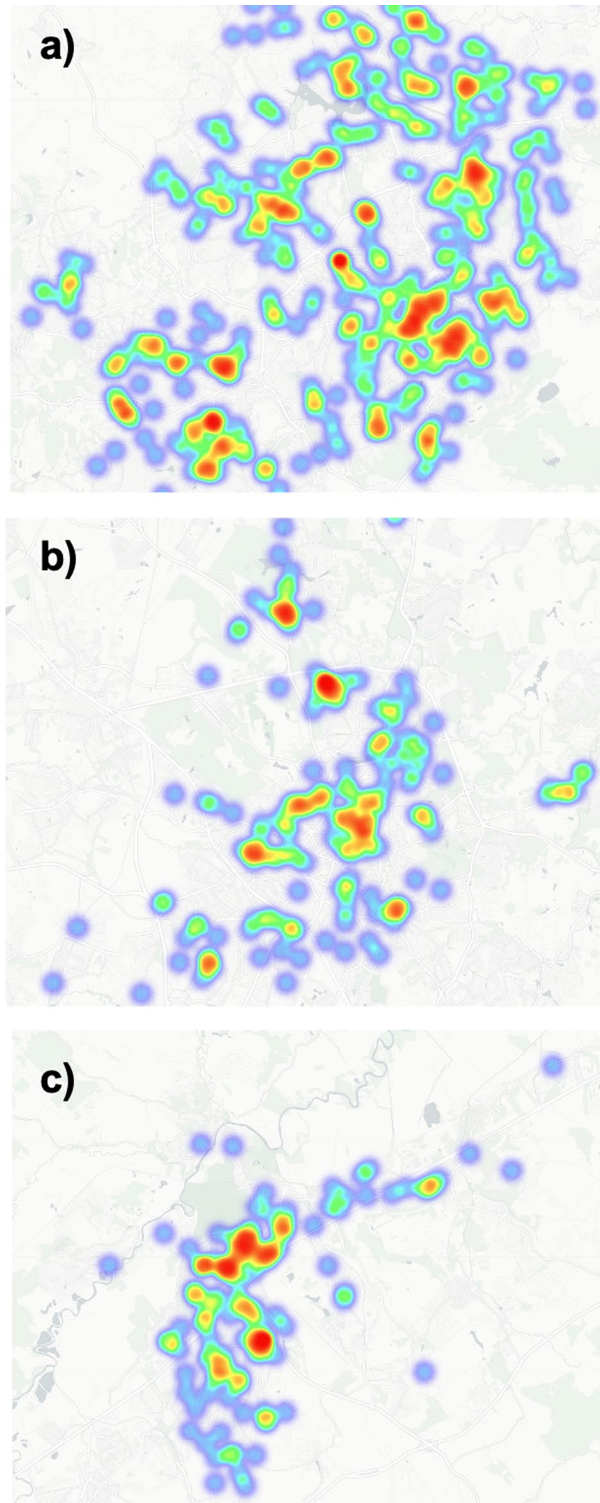

Fig. 2 Co-location mining process. Panel **a**, **b** and **c** represent the heat map of co-location positions of the drivers in cities of different sizes, with red implying a greater number of observations. The underlying map of the cities has been omitted to maintain the cities anonymous

Table 1 Summary of the cities used in the study

City	Population	Population coverage (%)	Area (km^2)	Region	Couriers	Distinct co-locations	Events
C2	2530701	87.4	274.0	Southeast	6278	41196	220630
C3	1963726	85.6	336.5	South	5102	18339	81906
C1	2703391	87.3	253.7	Northeast	18257	9843	41008
C4	1223237	73.4	245.1	Southeast	9659	11366	61508
C6	695328	76.3	134.9	Southeast	3312	5014	21296
C8	580870	80.7	129.4	South	702	6733	32258
C9	737310	60.9	128.9	Southeast	2233	7364	32862
C7	623614	75.3	160.6	Central West	1314	7575	26736
C5	871126	74.8	172.3	North	2625	5061	29286
C13	260690	73.6	74.4	Southeast	396	1628	6552
C12	282164	78.9	49.9	Southeast	476	1710	10098
C10	417478	79.4	73.5	Southeast	1751	1952	15748
C11	343643	70.8	77.5	Northeast	465	1456	10044
C14	95320	9.0	32.5	South	16	105	236
C15	11507	54.9	26.9	North	35	139	1248

Population refers to the total resident population of the city, coverage is the population coverage by the mobile phone application service. Area is the urban area of the city in km^2 (IBGE 2019). Region is the geographical region that groups several states. Couriers are the total quantity of active couriers in the city (with at least 1 order in the 14 days prior to or after the studied period). Distinct co-locations are the total count of unique latitude and longitude places where we have at least one co-location. Events are the total number of considered events after the data cleaning process. The chosen cities have a mean population density of 5176 ± 2402 residents per km^2

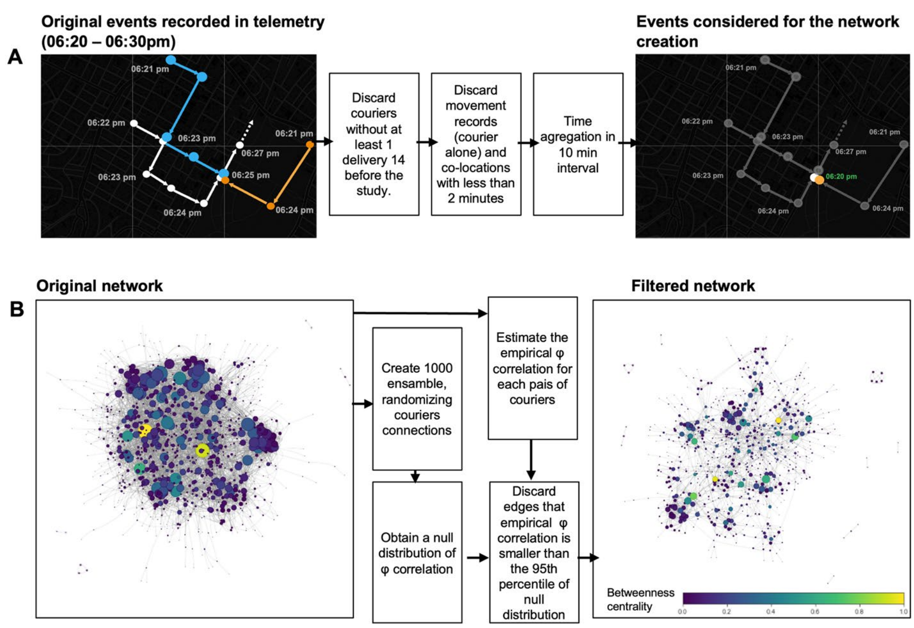

Fig. 3 Illustration of the steps employed in the preprocessing of the co-location data. Panel **a** represents a hypothetical scenario with 3 couriers (blue, white, and orange) that have a co-location at 06:26 pm, where just the events from couriers white and orange are filtered to create the network. Panel **b** displays the same network prior to and after filtering according to the steps described for city C9. The color represents the normalized betweenness centrality, and the size of the nodes is proportional to the degree of centrality

increasing observations lead to an increase or decrease in the number of co-locations between the two couriers, respectively.

We also filter links whose ϕ-correlation can be explained by pure chance alone. We generate a null distribution by bootstrapping the original network in each city to obtain an ensemble of $L = 1000$ randomly generated networks (Davidson 2007; Hardy and Mueser 2017; Snijders and Borgatti 1999). Each random network is generated by shuffling co-location events between pairs of couriers while maintaining the number of events associated with each courier (Guillaume and Latapy 2006).

In this manner, the process assumes that all generated networks are independent and identically distributed to the original network. For each randomization l, we calculate $\tilde{\phi}^l_{ij}$ associated with each courier pair, and the ensemble of such values form the null distribution $\tilde{\Phi}ij : \{\tilde{\phi}^1 ij, \tilde{\phi}^2_{ij}, \ldots, \tilde{\phi}^L_{ij}\}$. Using statistical inference methods (Gotelli 2000), we estimate the p-value of ϕ_{ij} by calculating the upper tail probability of obtaining a value equal to or greater than ϕ_{ij} from the cumulative frequency of the null distribution $\tilde{\Phi}_{ij}$. We discard links with a significance threshold of p value > 0.05.

Due to these steps, between 64 and 86% of edges and between 23% and 47% of nodes have been discarded from the original unfiltered networks. Figure 3B illustrates and summarizes the above-described steps and shows the network of courier co-locations obtained from city C9.

Results and discussion

We start by comparing the networks from the different cities. In particular, how the characteristics of these networks depend on the city size measured by the population covered by the service. Looking into cities of different dimensions and properties provides an opportunity to understand to which extent the scale of a city dictates the emergence of social networks with similar topological features. We focus our analysis on the giant component of each network.

Although these two cities are considerably smaller in size and population coverage, they provide observations at the lower-end of the population size range. In particular, they might indicate the existence of a critical city size above which courier networks emerge and become functional and below which they are absent. An idea that we believe can be worth future empirical or theoretical work (Table 2).

Also, in Fig. 4, we show the scaling relationship $D(k) \approx k^{\alpha}$, pointing to the possible heavy-tail degree distribution and an increasing degree heterogeneous character of the networks with increasing population size. There is no apparent convergence to a particular scaling coefficient with population size. Instead, we see a steady decrease approaching $\alpha = -2.0$, but all observations remain below that threshold.

Then, we look at the existence of other scaling relationships as a function of the size of the covered population. We show that, as expected, the average degree (Fig. 5A), maximum degree (Fig. 5D), and degree variance (Fig. 5B) grow with the population size. Moreover, larger cities lead to networks that are sparser (Fig. 5I) but also with a longer diameter (Fig. 5C) and greater average path length (Fig. 5E). In contrast, the clustering coefficient (Fig. 5F) is relatively stable, with an average of 0.20 ± 0.1 and a low R^2. Finally, we also note that the modularity (Fig. 5G) of these networks increases

Table 2 Summary of the statistics of the Giant Component of the couriers' informal social networks inferred from the co-location events

City	N	E	$\langle k \rangle$	σ^2	k_{max}	CC	APL	ρ	$Diam$	M
C2	3735	14484	7.76	77.53	55	0.27	5.73	0.36	16	0.84
C3	1868	5138	5.50	37.48	36	0.22	6.65	0.35	16	0.86
C1	1279	2679	4.19	19.28	33	0.16	6.42	0.32	16	0.81
C4	1071	3152	5.89	41.22	45	0.28	5.72	0.4	15	0.82
C9	714	1929	5.40	23.20	27	0.25	4.94	0.19	12	0.78
C8	680	1746	5.14	26.40	35	0.18	4.93	0.18	15	0.73
C6	651	1329	4.08	17.25	23	0.14	5.50	0.27	12	0.78
C7	490	1270	5.18	21.66	29	0.26	4.95	0.2	13	0.76
C5	437	1382	6.32	30.53	27	0.27	4.21	0.09	11	0.71
C12	164	361	4.40	13.26	19	0.23	4.62	0.23	13	0.66
C10	147	306	4.16	11.57	15	0.24	3.86	0.16	9	0.59
C13	138	199	2.88	4.63	11	0.13	4.78	0.07	10	0.70
C11	120	275	4.58	11.25	15	0.18	3.79	0.01	10	0.57
C14	8	5	1.25	0.18	2	0	1.33	− 0.67	2	0.64
C15	7	4	1.14	0.12	2	0	1.33	− 0.33	2	0.62

N is the total number of nodes, E is the total number of edges, $\langle k \rangle$ is the average degree, σ^2 is the variance of the degree, k_{max} is the maximum degree, CC is the cluster coefficient, APL is the average path length, ρ is the assortativity mixing, $Diam$ is the diameter and M is the modularity

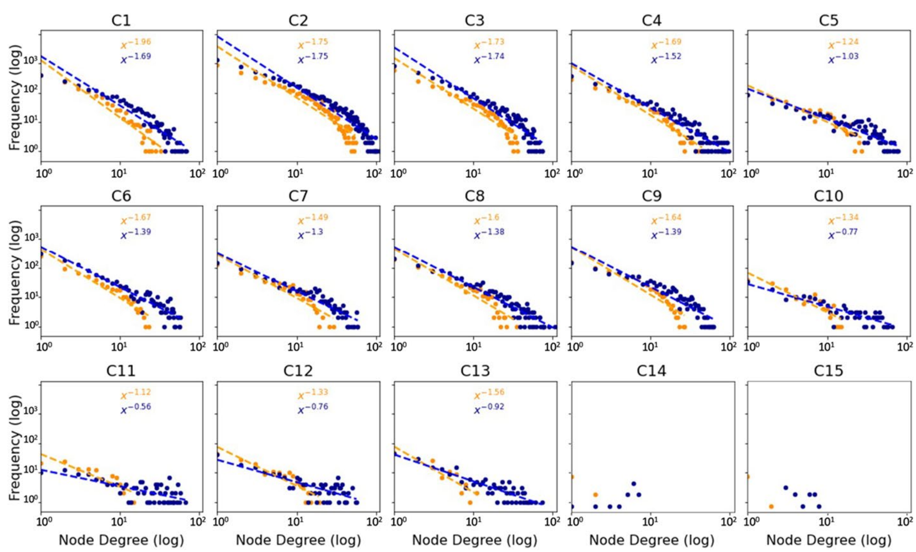

Fig. 4 Comparison of the Degree Distributions between the 15 cities of study. In blue is the degree distribution of the original co-location-based social networks. In orange, the degree distribution of the treated networks, filtered according to the steps discussed in the main text. The dashed line represents the best fit OLS linear model, and the domain of the line indicates the domain used for fitting the curve, which was truncated to remove the effect of the cut-off that is, in some cases, clearly visible

with population size, the same happens the degree-degree correlations (Fig. 5H) even though these remain relatively low.

The relatively high levels of modularity (mean of 0.63 ± 0.27) can raise interesting questions. One is that these networks can be modulated by the fact that local cities

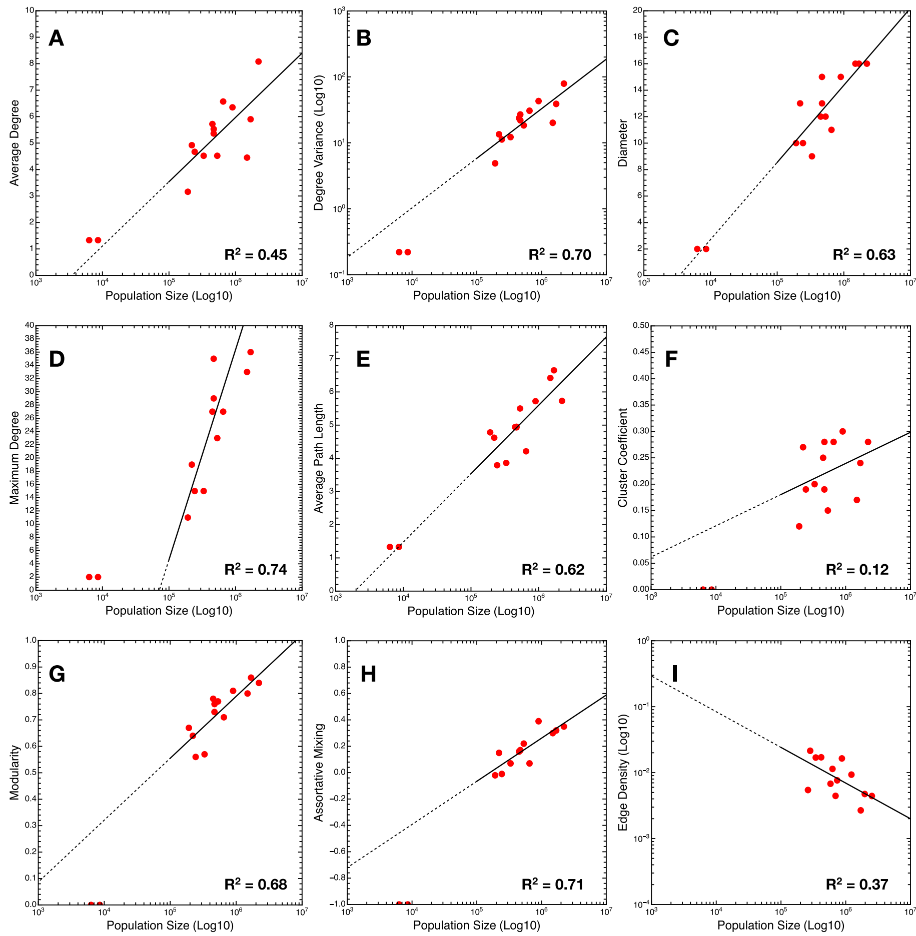

Fig. 5 Relationships between different network metrics and the population size of each city. Panels illustrate the results for the Average Degree (**A**), Log of the Variance of the degrees (**B**), diameter (**C**), Maximum degree (**D**), Average Path Length (**E**), Cluster Coefficient (**F**), Modularity (**G**), Assortativity Mixing (**H**), and Log Edge Density (**I**). Each red dot corresponds to observation for the cities C1 to C13, full lines correspond to the best OLS linear fit (R^2 values are displayed), and dashed lines show the extrapolation to cover the range of two smallest cities (C14 and C15)

reached their carrying capacity on the number of couriers necessary to perform the deliveries, resulting in the segregation of the couriers in different areas of the city and, thus, the specialization of couriers in specific locations. In other words, groups of couriers form compact communities that specialize in localized regions in the city and serve specific vendors (i.e., restaurants). This can be seen in Fig. 6, which shows the lack of overlap between different communities and the occupation frequency per location. A consequence of this finding is that such modularity and formation of communities can lead to cartel behavior between couriers and challenging for new couriers to operate in certain areas. Risks that become more prevalent in larger cities. Note that, importantly, in this analysis, we contrast the preferred locations of the network communities formed by co-location events.

Overall, and although we study cities spanning population sizes as much as three orders of magnitude, we do not find consistent evidence that the social network of couriers converges to a particular topology with stable features, as shown in Fig. 7. We find

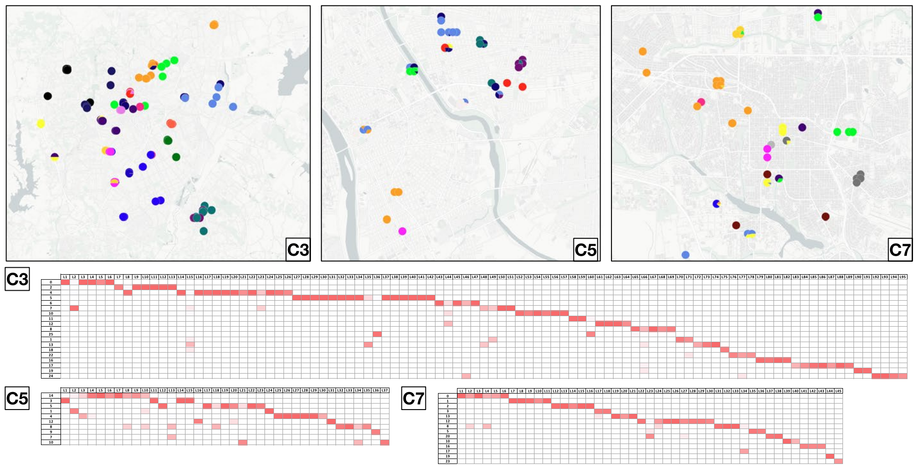

Fig. 6 Top panels show the locations in three cities—C3, C5, and C7—used in the co-location of couriers with colors showing the prevalence of couriers from different network communities in such locations. The bottom panels show the occupation matrices of communities (rows) per location (columns), showing a diagonal pattern indicating that communities do not often overlap between communities. Red indicates the high prevalence of a community in a location, white means no prevalence

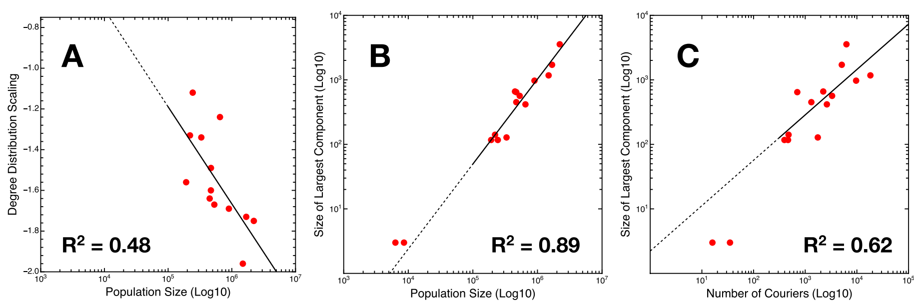

Fig. 7 Panel **A** shows how the scaling of the degree distribution depends on the city population size. Panel **B** shows how the size of the largest connected component of the network scales with the city population size. Panel **C** shows how the size of the largest connected component scales with the number of operative couriers in each city. Each red dot corresponds to observation for the cities C1 to C13, full lines correspond to the best OLS linear fit (R^2 values are displayed), and dashed lines show the extrapolation to cover the range of two smallest cities (C14 and C15)

structures with properties that scale alongside the city size and that fail to show the emergence of universal character. Instead, we show that these structures are possibly limited by the geographical activity patterns and organization/segmentation of couriers.

Going further, and as an example of the potential application of these networks, past works in social networks have shown a positive relation between node centrality and individual performance (Joksimović et al. 2016; Zhao 2022; Freeman 2002). In that sense, Degree centrality is commonly interpreted as a measure of the popularity of individuals (Mullen et al. 1991); the Betweenness centrality focuses on quantifying the role of individuals as brokers of information between distinct components; and the Eigenvector centrality emphasizes the influence or prestige of an individual based on their connections to other influential or high prestige individuals. In that sense, we look into the relationship between these three centrality measures with the income, as a performance proxy of couriers. Even though couriers can work long hours on the platform, there is no guarantee of income, and they need to be alert to choose a good place to wait for a delivery and choose wisely the time of the day they will work.

The top panels of Fig. 8 show the relationship between the above-described centrality measures and income. We identify a positive but small relationship between centrality and income. To statically validate this difference, we performed a one-way analysis of variance (ANOVA) to examine the differences in the degree centrality variable among the five categories (quintiles) of the independent variable for the filtered network.

The ANOVA results revealed that the independent variable had a significant effect on the degree centrality variable, $F(4, 6800) = 37.27, p\,value < 0.001$. However, as the degree centrality variable was not normally distributed (Shapiro–Wilk test, $W = 0.35, p\,value < 0.001$) and the assumption of homogeneity of variances was violated (Levene test, $F(4, 6800) = 17.98, p\,value < 0.001$), we chose a non-parametric test as a more robust alternative. Moreover, under the Kruskal–Wallis H-test, a robust test is suitable for comparing the medians of a continuous dependent variable across multiple independent groups—comparing the medians of the degree centrality across the quintiles (Mishra et al. 2019). The Kruskal–Wallis H-test revealed a significant difference between groups, $H(4) = 711.39, p\,value < 0.001$. To determine which pairs of quintiles were significantly different, we used a post hoc Dunn's test with Holm correction (Pereira et al. 2015). The median of the degree centrality was significantly higher in the second quintile than in the first, the third than in the second, the fourth than in the third, and finally, in the fifth than in the fourth, corroborating the literature findings where centrality and performance are positively associated.

Naturally, one can argue that such a positive relation between income and centrality can be due to the tenure of couriers. That is, older couriers had more time to choose the best waiting locations and thus achieve the most central positions in the social network.

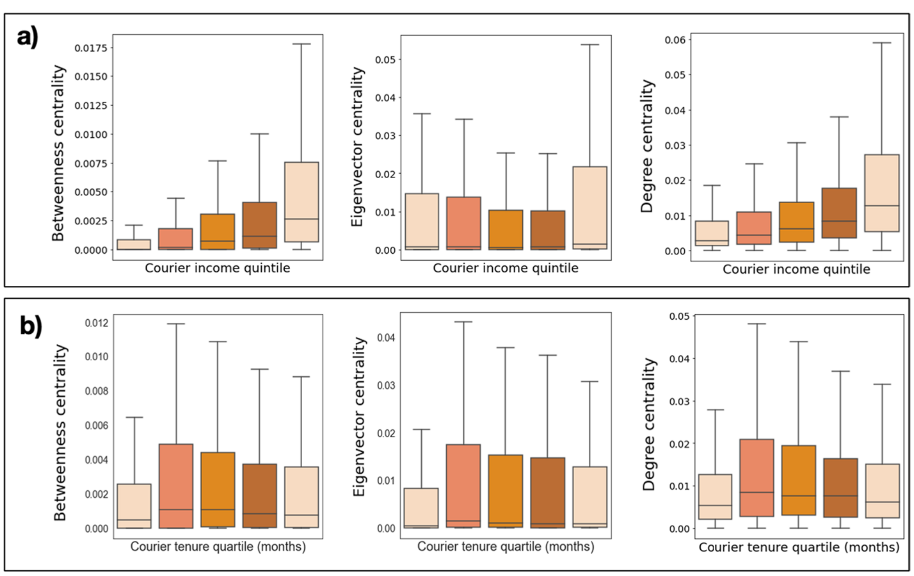

Fig. 8 Centrality measure according to income and tenure quintiles in filtered networks. Panel **a** shows the boxplots with the distribution of the betweenness centrality, eigenvector centrality, and degree centrality by income quintiles. In contrast, panel **b** shows the same centrality measures for tenure (time since first login in the delivery platform) quintiles

However, that is not the case, as the bottom panels of Fig. 8 show no relationship trend. Moreover, the same test validation process with the income was also done with the tenure (the time difference between the moment of the dataset retrieval and couriers' registration on the platform). Although there is some statically significant difference in the median among the quintiles, it happens just for some o them, and there is no correlation between tenure and degree centrality, as can be visually inferred from Panel b of Fig. 8. We found that the positive difference between the first and the second and between the third and the fourth quintiles are significant, while the difference between the second and the third is not. Further, we can identify a slight decrease in degree centrality value while tenure increases.

Both income and tenure values were omitted due to data privacy.

Final remarks

In this manuscript, we explored the potential of mining a co-location dataset of couriers collected from a food delivery platform in Brazil, comprised of 15 cities with different characteristics. By matching the couriers using co-location and filtering the dataset based on the significance of the ϕ-correlation coefficient, we have inferred a courier-courier network comprising 11,509 nodes and 34,295 unique edges. Furthermore, we differentiated coincidental co-locations from statistically significant relations in the location-based social network using the co-location data generated from the telemetry of couriers' mobile when working.

As practical implications of our findings, we can have two perspectives. On the one side, the courier's perspective of building a network can add voice to this class, creating a natural labor market organization that can challenge any aspect of work exploitation. In sum, the courier's network can express organic forms of collaboration between themselves (Wood et al. 2019). On the other side, from the delivery platform, it may be rich information highlighting the central players in the network, separating from the peripherical players, which could facilitate the communication between the platform and workers and improve actions such as retention of couriers (i.e., the central position of a courier enables him to obtain more significant support from their sibling). Also, by standard, the gig platforms do not provide tools to improve worker connections, inhibiting an opportunity for workers to collaborate among themselves and lower the education cost inherent to new couriers, reducing boundaries to the change of explicit and tacit knowledge. Identifying the network components with higher density may facilitate creating the courier profile and make it easier to detect deviant behaviors, such as fraudsters (Gao et al. 2019).

In summary, identifying gig workers' networks and understanding the relationship between centrality and performance can provide valuable insights for gig economy companies. These insights can be further explored, and here is a not exhaustive list of applications of these insights: optimize task allocation by allocating order to higher performance—higher centrality—couriers, promote collaboration, identify influencers, target training, and development, enhance worker engagement, and monitor performance. This can ultimately lead to a more efficient, satisfied, and high-performing workforce. Future research may address how these informal interactions highlighted in the co-location between couriers foster knowledge spread and the adoption of best practices, which can impact their performance, controlling for variables such as online hours, delivery distance, and other features that can affect couriers' behavior and location.

A limitation of this study is the lack of data at the courier level, which could have opened the door to a deeper exploration of the link between network features and individual performance or analytics. However, at the time of writing this manuscript, the platform did not liberate individual data, and as such, our results are preliminary and highlight potential interest for future research that is supported by the network analysis of the couriers' informal network. Further research using other performance metrics (i.e., customer evaluation, orders delivery per online time, among others) at the couriers' level can foster the knowledge about network formation and betweenness centrality impact as well as create a better understanding of informal interactions between couriers, fosters knowledge spread and the adoption of best practices, which can impact their performance and morale.

The ϕ-correlation coefficient only considers the similarity of interactions between two courier nodes without considering the specificity of the interactions. There are two main issues that it may cause. Firstly, the presence of location hubs may lead to artificially high levels of interaction profile similarity, as a large number of couriers may bind to a single location hub, leading to an interaction that is not very informative. Secondly, not all locations where the couriers may meet each other are independent. It may also lead to excessive levels of interaction similarity. For example, the courier's work time preference may lead to locations near restaurants with more probability of having an order for that time of the day (for example, meat at lunchtime and pizza during dinner). Also, using ϕ -correlation and co-location data, we are unable to differentiate between friendship and knowledge-seeking connections in the network (Yuan et al. 2011).

The data may contain bias regarding the region and time it was captured since it was available only one time period for some cities inside the same country and for the same food delivery platform.

Lastly, it is possible to argue that individuals who work long hours often engage in more interactions and earn higher incomes. Centralities tend to exhibit correlations, meaning that a high degree of centrality may be associated with other factors influencing income. While other factors may complicate this understanding (i.e., gig work income depends on market demand, task availability, operation patterns, skill level, efficiency, and compensation structure), analyzing these correlations in the present study was impossible. Further research may elaborate on these aspects and measure the influence of confounding factors

The process described in this article and the filtered networks form the boilerplate to study social influence and strategic diffusion dynamics across networks of collaborators in services of the gig economy. It is worth noting that the gig economy workers have moved from a side job executed as an income complement to a full-time position with a flexible schedule, which brings a rich amount of research opportunities to understand how it affects the performance and the network itself—a hypothesis is that two communities are created: a part-time and a full time. We hope this article sheds some light on these gig workers and that more research is conducted to deepen the knowledge about this new labor market.

Acknowledgements
GP, AM, and FLP are very grateful for the suggestions given by the audience and peer review of the Complex Networks and Their Applications XI conference, in which we were able to clarify some points and enrich the research. The authors are thankful to the food delivery platform for sharing the data for this study. The findings, interpretations, and conclusions expressed by the authors in this work do not necessarily reflect the views of the food delivery platform.

Author contributions
All the authors discussed and designed the proposed research; GP acquired the data, wrote the code, executed the experiments, and analyzed and interpreted the courier data. FLP and CC interpreted the statistics, supervised the study, and conceived the idea. All authors participated in the writing of the manuscript. Finally, all authors read and approved the final manuscript.

Funding
FLP acknowledges the financial support provided by FCT Portugal under the project UIDB/04152/2020 – Centro de Investigação em Gestão de Informação (MagIC).

Availability of data and materials
The data supporting this study's findings are available from the food delivery company, and restrictions apply to the availability of these data, which were used under license for the current research and are not publicly available. Data are, however, available from the authors upon reasonable request and with the explicit permission of the food delivery company.

Declarations

Competing interests
The authors declare that they have no competing interests.

Received: 27 March 2023 Accepted: 22 October 2023

References

Candia C, Encarnação S, Pinheiro F (2019) The higher education space: connecting degree programs from individuals' choices. EPJ Data Sci 8:39
Candia C, Maldonado-Trapp A, Lobos K, Peña F, Bruna C (2022) Disadvantaged students increase their academic performance through collective intelligence exposure in emergency remote learning due to covid 19. arXiv preprint arXiv:2203.05621
Cherry MA (2016) People analytics and invisible labor. Louis ULJ 61:1
Crandall DJ, Backstrom L, Cosley D, Suri S, Huttenlocher D, Kleinberg J (2010) Inferring social ties from geographic coincidences. Proc Natl Acad Sci 107(52):22436–22441
Davidson R (2007) Bootstrapping econometric models. Technical report, McGill University, Department of Economics
Evans PC, Gawer A (2016) The rise of the platform enterprise: a global survey. Technical report, The Center for Global Enterprise
Freeman LC et al (2002) Centrality in social networks: conceptual clarification. Soc Netw Crit Concepts Sociol 1:238–263
Gao F, Wang J, Wang S (2019) Understanding knowledge workers' job performance: a perspective of online and offline communication networks. Enterp Inf Syst 13(1):107–131
Gotelli NJ (2000) Null model analysis of species co-occurrence patterns. Ecology 81(9):2606–2621
Guillaume JL, Latapy M (2006) Bipartite graphs as models of complex networks. Physica A 371(2):795–813
Gupte M, Eliassi-Rad T (2012) Measuring tie strength in implicit social networks. In: Proceedings of the 4th annual ACM web science conference, pp 109–118
Hardy KV, Mueser KT (2017) Trauma, psychosis and posttraumatic stress disorder
Hsieh HP, Yan R, Li CT (2015) Where you go reveals who you know: Analyzing social ties from millions of footprints. In: Proceedings of the 24th ACM international on conference on information and knowledge management, pp 1839–1842
Huang K, Yao J, Yin M (2019) Understanding the skill provision in gig economy from a network perspective: a case study of Fiverr. Proc ACM Hum-Comput Interact 3(CSCW):1–23
IBGE (2019) Tabela 8418 - Áreas urbanizadas, loteamento vazio, Área total mapeada e subcategorias. https://sidra.ibge.gov.br/tabela/8418#/n6/all/v/12749/p/all/d/v12749%204/l/v,p,t/resultado. Accessed 01 Mar 2023
Iyengar R, Van den Bulte C, Choi J (2011) Distinguishing between drivers of social contagion: insights from combining social network and co-location data. Dartmouth College, Hanover, NH, Scheduled
Joksimović S, Manataki A, Gašević D, Dawson S, Kovanović V, De Kereki IF (2016) Translating network position into performance: Importance of centrality in different network configurations. In: Proceedings of the sixth international conference on learning analytics & knowledge, pp 314–323
Kalgotra P, Sharda R, Luse A (2020) Which similarity measure to use in network analysis: Impact of sample size on phi correlation coefficient and Ochiai index. Int J Inf Manag 55:102229
Kinder E, Jarrahi MH, Sutherland W (2019) Gig platforms, tensions, alliances and ecosystems: an actor-network perspective. Proc ACM Hum-Comput Interact 3(CSCW):1–26
Lu S, Zhao J, Wang H (2022) Academic failures and co-location social networks in campus. EPJ Data Sci 11(1):10. https://doi.org/10.1140/epjds/s13688-022-00322-0

Md Fadzil FN, Che Azmi A (2022) Establishing factors affecting the tax morale of individuals working in the gig economy. J Glob Responsib 13(2):157–176
Mishra P, Pandey CM, Singh U, Keshri A, Sabaretnam M (2019) Selection of appropriate statistical methods for data analysis. Ann Card Anaesth 22(3):297
Mullen B, Johnson C, Salas E (1991) Effects of communication network structure: components of positional centrality. Soc Netw 13(2):169–185
Njoo GS, Kao MC, Hsu KW, Peng WC (2017) Exploring check-in data to infer social ties in location based social networks. In: Advances in knowledge discovery and data mining: 21st Pacific-Asia conference, PAKDD 2017, Jeju, South Korea, May 23–26, 2017, proceedings, part I, vol 21. Springer, pp 460–471
Njoo GS, Hsu KW, Peng WC (2018) Distinguishing friends from strangers in location-based social networks using co-location. Pervasive Mob Comput 50:114–123
Pereira DG, Afonso A, Medeiros FM (2015) Overview of Friedman's test and post-hoc analysis. Commun Stat Simul Comput 44(10):2636–2653
Perren R, Kozinets R (2018) Lateral exchange markets: how social platforms operate in a networked economy. J Mark 82(1):20–36
Pi T, Cao L, Lv P, Ye Z, Wang H (2018) Inferring implicit social ties in mobile social networks. In: 2018 IEEE wireless communications and networking conference (WCNC). IEEE, pp 1–6
Pilatti G, Pinheiro FL, Montini A (2023) Gig economy and social network analysis: Topology of inferred network. In: Complex networks and their applications XI: proceedings of the eleventh international conference on complex networks and their applications: complex networks 2022, vol 2. Springer, pp 471–479
Psorakis I, Roberts SJ, Rezek I, Sheldon BC (2012) Inferring social network structure in ecological systems from spatio-temporal data streams. J R Soc Interface 9(76):3055–3066
Ronen S, Gonçalves B, Hu KZ, Vespignani A, Pinker S, Hidalgo CA (2014) Links that speak: the global language network and its association with global fame. Proc Natl Acad Sci USA 111:E5616–E5622
Snijders TA, Borgatti SP et al (1999) Non-parametric standard errors and tests for network statistics. Connections 22(2):161–170
Tubaro P (2021) Disembedded or deeply embedded? a multi-level network analysis of online labour platforms. Sociology 55(5):927–944
Vallas S, Schor JB (2020) What do platforms do? understanding the gig economy. Ann Rev Sociol 46:273–294
Weidenstedt L, Geissinger A, Leick B, Nazeer N (2023) Betwixt and between: triple liminality and liminal agency in the Swedish gig economy. Environ Plan A Econ Space, 0308518X231172984
Wood AJ, Graham M, Lehdonvirta V, Hjorth I (2019) Networked but commodified: the (dis) embeddedness of digital labour in the gig economy. Sociology 53(5):931–950
Woodcock J, Graham M (2019) The gig economy. A critical introduction. Polity, Cambridge
Yeganeh H (2021) An analysis of factors and conditions pertaining to the rise of the sharing economy. World J Entrep Manag Sustain Dev 17(3):582–600
Yuan YC, Rickard LN, Xia L, Scherer C (2011) The interplay between interpersonal and electronic resources in knowledge seeking among co-located and distributed employees. J Am Soc Inform Sci Technol 62(3):535–549
Zhao J (2022) Coupling open innovation: network position, knowledge integration ability, and innovation performance. J Knowl Econ 14:1–21

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.