



Self-loops in Social Networks: Behavior of Eigenvector Centrality

J. J. Merelo¹ and M. Cristina Molinari²

¹ Department of Computer Engineering, Automatics and Robotics and CITIC
University of Granada, Granada, Spain

jmerelo@ugr.es

² Dipartimento di Economia, Ca' Foscari University of Venice, Venice, Italy
cmolinar@unive.it

Abstract. Centrality measures are an essential tool in understanding complex networks, since they give researcher insights on the role the different nodes/actors play in them. Among them, eigenvector centrality is a principled approach to these measures, using a mathematical operation on the connection matrix. This connection matrix includes connections from an actor to itself (the diagonal); however, as it is the case with most centrality measures, this fact is seldom used in social studies to compute the standing or influence of one node over others. In this paper we will analyze the difference in EV centrality with or without these self connections or self-loops and how the change depends on the actual number of these self-loops or the weight of these self-connections. Finally, we will characterize in which cases, if any, it is effective to drop self-loops and what kind of information it will give us on the nature and dynamics of the network.

[AQ1](#)

[AQ2](#)

Keywords: Complex networks · Social networks · Graph Theory · Eigenvector centrality

1 Introduction

Eigenvector centrality [1] measures how an actor in a network or graph influences other actors by computing the eigenvalues of the adjacency or connectivity matrix, that has as components the weight of the connection of every agent to all the others, including itself. It is a centrality measure because, along with other such measures (such as betweenness or degree centrality [17]), it is a micro-level assessment of the power or influence of a node within a network through the analysis of its connections. It has been used extensively in social network analysis [5, 12, 14]; but also in other fields, such as biology [9], economics [15], or even in the analysis of the spread of opinions in internet forums during the COVID-19 pandemic [11].

In its widespread use for the computation of the value for a single node it is not very different from other centrality measures; however, unlike them,

it considers the whole network [2] taking into account all weights (including negative ones); and, more importantly for the purposes of this paper, all weights include the connections of a node to itself, or self-loops.

As a matter of fact, many complex networks, including social networks, include links from one actor/agent to itself; we can find them in commuting networks [8], where self-loops would indicate trips that start and end in the same city, county or state; in family networks [8], with self-loops indicating marriage between members of the same (extended) family; opinion networks [11], where they would indicate interaction among members of the same group; sport teams transfer networks [7], indicating transfers between teams in the same country; when freight traffic networks [10] are analyzed at a regional level, self-loops would represent shipments that start and end in different parts of the region; commercial networks [15] where self-loops indicate a contract between members of the same family, or even biological networks [9], where self-loops could indicate auto-trophic (members of a species feeding on other members) or auto-catalytic (reactive agents catalyzing reactions where they participate).

The main issue and the one we are trying to address in this paper, is that when centrality analysis is performed on these networks, most centrality measures cannot work with self-loops; thus, in most cases, they are dropped [11, 14, 15]. Even when a specific measure can include self-loops, like the aforementioned eigenvector centrality that works on the connection matrix, self-loops are usually dropped mainly to work on an uniform set of data, that is, not use two different connection matrices with diagonal values (for EV centrality) and other without (for other measures).

In the cases where self-loops do not have a clear interpretation (or at least a common interpretation with the rest of the connections) there are good reasons to work that way: betweenness centrality [6], for instance, models how one node is needed to transmit information to other parts of the network; how often it transmits information to itself cannot easily be accommodated within this framework, and it can be argued that it could be simply ignored when computing the geodesics from which this measure is computed. That is not the case for EV centrality, which does in fact use self-loops to compute the eigenvalues that are presented as the EV centrality measure. And these intra-links can, however, give us better insights on the dynamics of a social network, and should probably be taken into account. So far, however, there has been little analysis on what is the actual effect of considering these intra-links for computing the eigenvector centrality [2] in social networks. This paper will try, using well-chosen examples, what is that effect and how it could help to better understand social systems.

The main research question that we ask, then, is if self-loops should be included, when available, in the computation of eigenvector centrality. As an accessory question, we try to investigate what would be the effect of doing so in EV centrality measurements, and how it affects the value and the ranking of nodes in the corresponding network.

The rest of the paper is organized as follows: next we will be describing the state of the art, to proceed to describe the datasets and the experiments

performed in Sect. 3. We will then discuss the results and conclude with our conclusions and future lines of work.

2 A Brief Literature Survey

The main motivation behind this paper is to shed a bit of light on the use of self-loops in social network analysis, mainly because it has been used so rarely in the past. Some papers acknowledge its importance in the propagation of information in social networks [22], with self-loops representing simply re-posts of some content previously created by the same person; other papers, however [11] dealing with the same subject, explicitly do not use them, thus missing a good amount of the dynamics created by these self loops. In other opinion formation papers [19] self-loops represent the amount of attention a person pays to its own opinion as opposed to others. In general, even theoretical models of the spread of information in social networks [3] include self-loops.

In general we can say that while self-loops are sometimes acknowledged and used in social network *models*, they are dropped when making centrality analysis of complex or social networks. This why we have made it the main focus of this paper.

3 Experiments

We will be using two datasets to perform the experiments. The first one is the Venetian matrimonial dataset [15], a social network of the noble families of the Republic of Venice from the XIV to the XIX century¹; the second is a dataset of freight traffic among the states of the United States of America extracted from the government data portal [20]. These will be examined in turn in the next two subsections.

3.1 Analyzing the Venetian Matrimonial Dataset

This dataset was published by Puga and Treffler as support for their paper [15]. It consists of marriages celebrated in the Republic of Venice (and successor polities during the late XVIII and XIX century) where the groom is a noble², registered with the *Avvogaria del Comune* of the Republic. Families (called *casate* in the original Venetian and Italian) were the political and social unit in the republic

¹ Some marriages are not dated, but we can assume they took place in the same range of years.

² Since the dataset includes some marriages that happened after the fall of the Republic in 1796, the concept of “noble” in this case corresponds to families that were considered noble *during the existence of the Republic*; during French and Austrian control, as well as during the brief period of the Republic of San Marco, such nobility titles no longer had any value; however, since they were included in the original dataset there was no good reason to eliminate them.

[16, 18], which explains the interest this dataset has for insights into the history of the republic. Nodes in this dataset are, thus, families with nobility titles in the Republic of Venice; we eliminated from this dataset those marriages where brides did not belong to any patrician family.³

This dataset has been chosen because it includes a good number of intra-family marriages, as well as because it has been studied extensively by Puga and Treffer in their paper, providing a basis for the comparison of results.

Table 1. Intra-family marriages per century

Century	Intra-family marriages	Total marriages	Percentage
XIV	5	51	9.80
XV	82	2740	2.99
XVI	118	4265	2.77
XVII	104	2770	3.75
XVIII	31	1732	1.79
XIX	3	108	2.78
NA	42	561	7.49

Table 2. Ranking of top ten families according to number of intra-family marriages (left).

Family	#Intra-family marriages
Contarini	65
Morosini	23
Corner	19
Martinengo	16
Querini	15
Balbi	12
Donato	11
Malipiero	10
Zorzi	10
Zancaruol	9

The absolute and relative number of intra-family marriages in each century is shown in Table 1, including marriages whose date is unknown. These marriages

³ There were many restrictions to this kind of marriage, but they occurred with regularity, at least until the so-called “Second Serrata” [4], during the XV century; in this case, however, we eliminate them because they are irrelevant to the main point of the paper, not having any influence in the EV centrality of a specific node.

are not evenly distributed per family; the top 10 families according to number of intra-family marriages is shown in Table 2. This is an example where the number of self-loops is not excessive, although it is certainly significant, hovering around a few percentage points per century for an amount of 385 intra-family marriages (self-loops) over 12227 total.

Table 3. Ranking of top ten families according to EV centrality values, with (right) or without (left) self-loops.

Family	EV	Family	EV + self-loops
Contarini	1.0000000	Contarini	1.0000000
Morosini	0.8175570	Morosini	0.6827832
Corner	0.6425724	Corner	0.5131463
Querini	0.5045249	Querini	0.3852909
Priuli	0.4286878	Priuli	0.3139764
Dolfin	0.3771834	Giustinian	0.2806501
Giustinian	0.3719008	Dolfin	0.2775836
Michiel	0.3692852	Michiel	0.2743379
Zorzi	0.3629267	Zorzi	0.2723025
Loredan	0.3593036	Pisani	0.2649323

This dataset has been transformed into two different graphs; eliminating self-loops in one and leaving them in the other. In both cases we use undirected edges joining the two families of the partners in every marriage; the edge is weighted with the number of marriages between the families in the nodes; eigenvector centrality has then been computed for the two resulting graphs. A ranking of the top families according to their EV centrality is shown in Table 3.

Looking at Table 2 together with the left hand side of Table 3, we can see that there seems to be some correlation between the number of intra-family marriages and the EV centrality, even when we do not include it in the computation. Six out of ten families are the same, and the first five: Contarini, Morosini, Corner, Querini and Priuli, also appear in the same order. This might indicate either a common cause for both rankings (size of the family, for instance) or a cause-effect, or even a combination of the two: a family gets bigger since it is wealthy, and is wealthy due to its social capital; high EV centrality implies a lot of influence, and this begets wealth, that literally increases the fitness of the family making it big enough that intra-family marriages become possible and even common. This discussion, however, is beside the main point of this paper, although it should be noted that, in a way, self-loops are *factored in* in this specific case since intra-family marriages tend to appear more frequently in families with high EV centrality.

Including self-loops in the computation we see that there are small, but significant, variations: The Dolphins and Giustinians change their order in the ranking,

plus the Loredan family is dropped and substituted by the Pisani family⁴. It can also be seen that the difference between the first and second family in the ranking has doubled, and that, in general, the value of the normalized EV centrality has also decreased; since the EV values are normalized, this simply indicates that the difference between Contarini and the other families has increased, a fact in which, of course, self-loops have had a decisive influence.

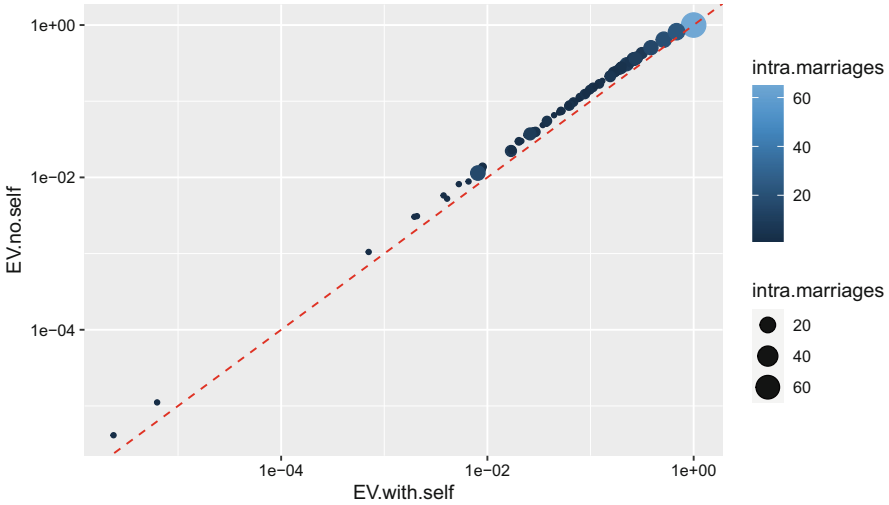


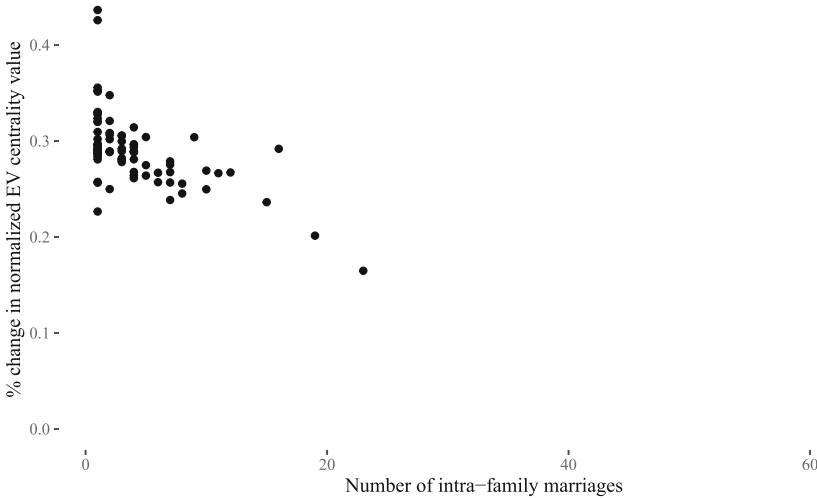
Fig. 1. EV centrality considering self-loops (x axis) or not (y axis). The size of the dot and color signal the number of intra-family marriages.

To better highlight the impact of self-loops, in Fig. 1 we have analyzed the differences between the EV values with (x axis) or without self-loops (y axis) for every node; it plots the values with self-loops (x axis) vs. those without (y axis) in a log scale. The plot shows that all values are over the $x = y$ line. The size and color of the dot representing every family is related to the number of intra-marriages; big, lighter blue dots will have the bigger number of self-loops, small, darker ones the lowest.

All dots are placed over the $x = y$ line, indicating that normalized EV values are lower without self-loops. Besides, looking at the colors and sizes, we can see that the smaller the size (number of marriages), the bigger the increase in value; that is, more intra-family marriages make the inclusion of self-loops decrease EV centrality *less*. However, this is due in this case to the fact that what is actually increased is the difference between the family with the higher EV (the Contarinis) and the rest, so we can look at this result from the other side: the

⁴ The Pisani family is certainly more “central” than the Loredan, at least looking at the number of nobles in important offices; [15] mentions them as one of the family with the greatest amount of shares in shipping contracts.

presence of a family with a high value in the diagonal of the connection matrix (i.e. self-loops with a high weight) increases its EV centrality much more than that of the rest of the nodes/families.



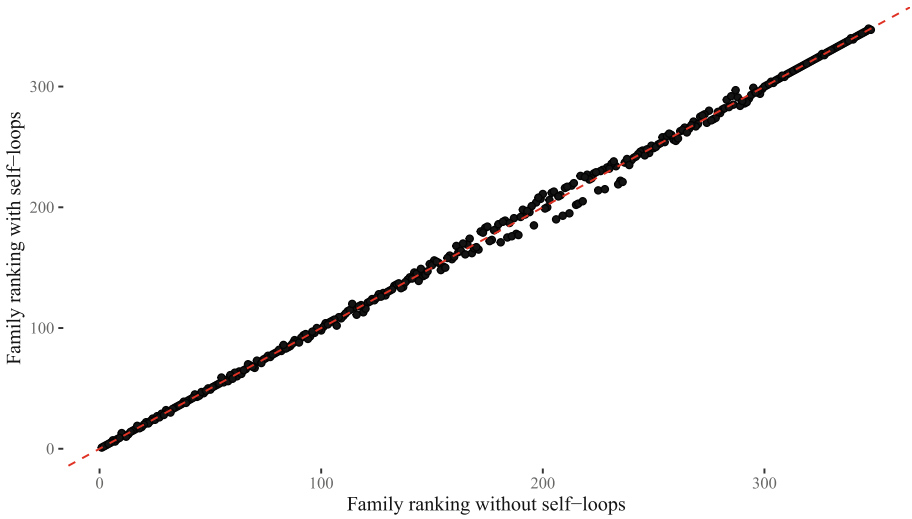


Fig. 3. EV centrality rank reached by families with (y axis) or without (x axis) self loops.

Table 4. Summary of linear fit of EV centrality for families in the XIV and XV century.

	RSE	R2	adjR2	Fstat	intercept.t	slope.t
With self-loops	0.056	0.825	0.823	733.504	7.306	27.083
Without self-loops	0.065	0.834	0.833	785.478	6.632	28.026

The influence of using self-loops (shown in the right-hand side panel) is clear, with points representing families with the highest EV seeming *closer* to the fit; it should be noted that these are the families that have a high-number of intra-family marriages. To quantify numerically the difference in Table 4 we show a summary of the coefficients of the regression model for the two data sets. The RSE column shows the difference of residuals between the model and the data; the top row is smaller, showing a better fit for the EV centrality values if self-loops are taken into account. This improvement is due mainly, looking at the two right-most columns, to the improvement in the fit of the intercept with 0 (the t value is better); the effect of this can be observed in Fig. 4, which shows how the red line that represents the model seems a bit more centered than the one on the left. The values of R-squared and its adjusted value, as well as the F-value, are slightly better for the model without self-loops; however, they are very similar and very high in both cases, so this difference is not considered significant.

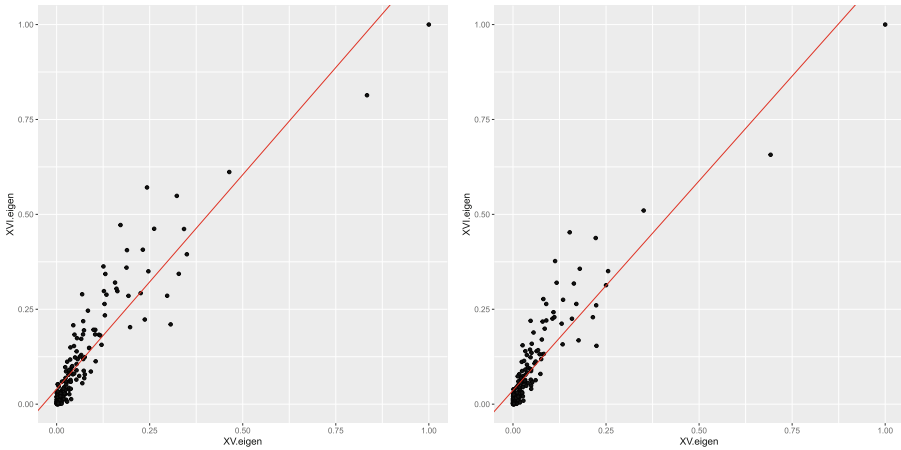


Fig. 4. Correlation between the EV centrality of Venetian families in the XV century (x) and XVI century (y). The line indicates the fitted linear model. Left-hand side chart does not include self-loops, right-hand side does.

At any rate, and to the extent that EV centrality in a century is a predictor for EV centrality in the next century, including self-loops in its computation makes it a better fit, even if the two models represent the data very well in both cases.

3.2 Freight Traffic Network

The freight traffic network is a directed network that includes as nodes different US states, and as edges the amount of freight traffic between them. The data is available from the US Bureau of Transportation Statistics on its page “Freight Flow by State”⁵. The data is available for the years 2017–2021.

We will be doing some additional processing on this data. All types of traffic (import, export, domestic) will be added into a single traffic flow; this quantity will be used as a weight in the transportation network. Also, we will use the data for the year 2021 only; the resulting processed data set is available from the GitHub repository for this paper <https://github.com/JJ/redes-venecia>. The resulting network is shown in Fig. 5. This network is totally different from the one analyzed in the previous subsection: all states have internal traffic, so all of them have self-loops, as the Figure shows. It is a directed network, since traffic between two states can be asymmetric. The network is very dense, with almost all states connected with all others. Finally, the network structure is quite different, with big differences between the most central states and the rest. Using self loops or not is bound to influence the vision we have of the centrality of the states.

⁵ <https://www.bts.gov/browse-statistical-products-and-data/state-transportation-statistics/freight-flows-state>.

Table 5. Ranking of states in the US freight network according to EV centrality values, with (right) or without (left) self-loops.

State	EV	State	EV + self-loops
Texas	1.0000000	Texas	1.0000000
California	0.9908871	California	0.4624587
Illinois	0.8105954	Illinois	0.0518211
New York	0.7649173	Michigan	0.0508070
New Jersey	0.6937505	Louisiana	0.0472489
Pennsylvania	0.6931582	Florida	0.0350895
Ohio	0.6796268	Ohio	0.0347809
Michigan	0.6607562	New York	0.0340062
Indiana	0.5740545	Tennessee	0.0298025
Tennessee	0.5117970	Georgia	0.0286042

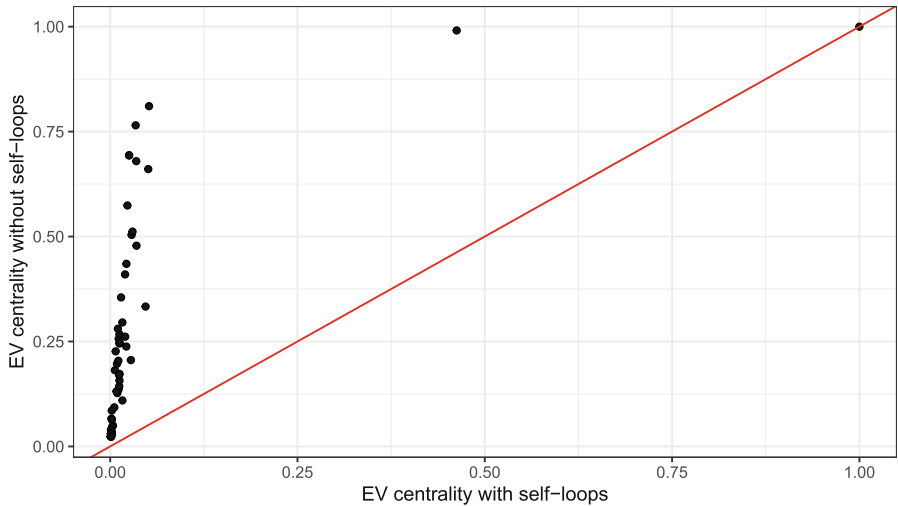


Fig. 6. Relationship between EV centrality with or without self-loops for the US freight network.

EV centrality, and also in general a higher self-connection weight will imply less change. This clearly indicates a correlation between the two measures, EV centrality and intra-connection weight, but it is not clear if it is a causal relationship or not, and discussing it is beyond the point of the paper.

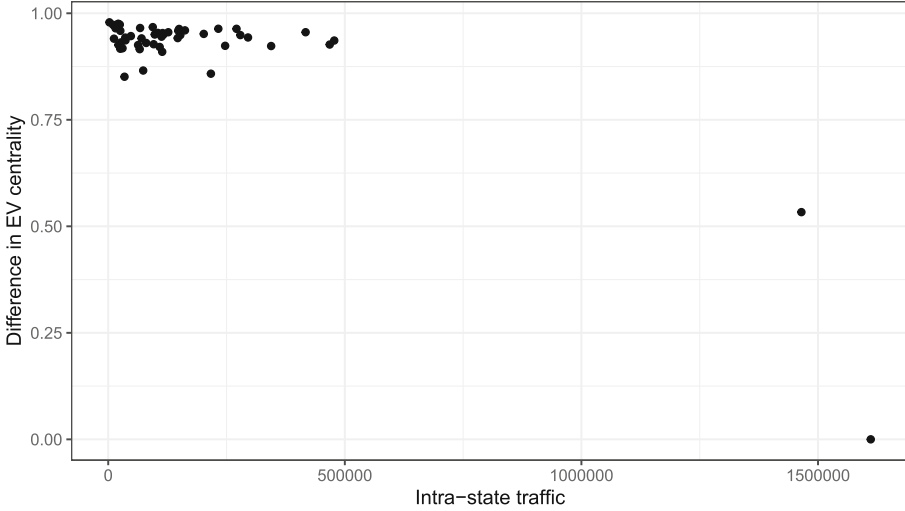


Fig. 7. Change in EV centrality vs. value of intra-state traffic.

4 Conclusions

The main research question we asked in the introduction was whether it was worth the while to take into account self-loops whenever possible in the computation of EV centrality. Using real-world networks, we have shown that, in general, the answer is yes. In the case of the Venetian matrimonial network it helps EV become a better predictor of the social status of the families involved, and in the case of the US freight network highlights the differences between the position in the network of states, boosting the difference between Texas, California, Illinois and the rest of the states, and boosting Michigan and Louisiana over New York and New Jersey given the importance of internal freight in these two states, that is, the weight of self-connections (diagonal values in the connection matrix). In both cases, the differences are not only quantitative, but also qualitative, as the ranking of the most important nodes changes when self-loops are included.

The direction and quantity of those changes is very similar in both networks examined. Complex networks reach a state that is the consequence of many different internal and external processes, and in general, the same processes that make a node in the network reach a high status will produce a high number of internal connections, that is, self-loops with a high internal weight. This makes the inclusion of self-loops *increase* the differences in status between high-ranked members, but also produce changes in ranking that are more pronounced in the case of mid-ranking members, as observed in the case of the Venetian marital network.

At any rate, the fact that we have observed very similar phenomena in two very different networks (sparse and dense, directed and undirected, with a low or high self-connection rate) probably indicates that the inclusion of self-loops in EV computation has greater importance than it has traditionally been given, and thus the recommendation to include self-loops in the computation of EV centrality should hold in general.

A immediate future line of work would of course try to test combinations of networks, possibly including synthetic ones, with different combinations of the above mentioned properties to actually characterize when and if the inclusion of self-loops is essential, merely interesting, or simply irrelevant. It can be argued that whenever they appear in a dataset they should be used, but to the extent that the physical interpretation of self-loops and how their data is collected is fundamentally different from external connections, including them or not is clearly an issue that should be researched and clarified in a case by case basis.

This probably indicates the need for inclusion of self-loops in other centrality measures, such as page-rank or betweenness centrality, would help us understand better social networks where those measurements explain better its dynamics; however, this will need a modification of the algorithms used to compute them, which is left as a future line of work. In order to be comprehensive, too, it would be interesting to see the relationship between the slope of the model shown in Fig. 1 and the relationship between the number of self-loops for all actors involved. Finally, investigating other social networks with the same characteristics would help us generalize these results and thus recommend to *never* eliminate self-loops in the analysis of social networks where they exist.

The most important line of work, however, is to check against some external measurement (number and importance of positions reached by a family, for instance, in line with [21]) which version of the EV centrality would be a better match, thus proving that self-loops/diagonal values in the connection matrix go beyond mere mathematical artifacts to have a precise and grounded interpretation; once that is proved, it would pave the way to an extension of the employment of self-loops in some way in other centrality values such as betweenness centrality.

Finally, we should note that this paper and the data used in it have been developed following the principles of Agile science [13] and its development can be observed in the repository <https://github.com/JJ/redes-venecia> together with the data used in it, making it fully reproducible using the same software. It is of course free software, and released under the terms of the GNU General Public License v3.0.

Acknowledgements. This work is supported by the Ministerio español de Economía y Competitividad (Spanish Ministry of Competitivity and Economy) under project PID2020-115570GB-C22 (DemocratAI::UGR).

References

1. Bonacich, P.: Factoring and weighting approaches to status scores and clique identification. *J. Math. Sociol.* **2**(1), 113–120 (1972). <https://doi.org/10.1080/0022250X.1972.9989806>
2. Bonacich, P.: Some unique properties of eigenvector centrality. *Soc. Netw.* **29**(4), 555–564 (2007)
3. Chierichetti, F., Lattanzi, S., Panconesi, A.: Rumor spreading in social networks. *Theoret. Comput. Sci.* **412**(24), 2602–2610 (2011)
4. Chojnacki, S.: Social identity in renaissance Venice: the second Serrata. *Renaissance Stud.* **8**(4), 341–358 (1994). <https://www.jstor.org/stable/24411934>
5. Cotta, C., Mora, A.M., Merelo-Molina, C., Guervós, J.J.M.: FIFA world cup 2010: a network analysis of the champion team play. *CoRR abs/1108.0261* (2011)
6. Freeman, L.C.: A set of measures of centrality based on betweenness. *Sociometry* 35–41 (1977)
7. GürsakaL, N., Sevilmiş, A., Aksan, A., Yılmaz, F.: Comparison of country-based high and low market valued transfer networks. *J. Curr. Res. Soc. Sci.* **10**(2), 417–430 (2020)
8. He, M., Glasser, J., Pritchard, N., Bhamidi, S., Kaza, N.: Demarcating geographic regions using community detection in commuting networks with significant self-loops. *PLoS ONE* **15**(4), e0230941 (2020)
9. Horstmeyer, L., Pham, T.M., Korbelt, J., Thurner, S.: Predicting collapse of adaptive networked systems without knowing the network. *Sci. Rep.* **10**(1), 1223 (2020)
10. Iyengar, D., Rao, S., Goldsby, T.J.: The power and centrality of the transportation and warehousing sector within the US economy: a longitudinal exploration using social network analysis. *Transp. J.* **51**(4), 373–398 (2012)
11. Kuikka, V.: Opinion formation on social networks - the effects of recurrent and circular influence. *Computation* **11**(5) (2023). <https://doi.org/10.3390/computation11050103>, <https://www.mdpi.com/2079-3197/11/5/103>
12. Merelo, J., Cotta, C.: Who is the best connected EC researcher? Centrality analysis of the complex network of authors in evolutionary computation. *Arxiv preprint arXiv:0708.2021* (2007)
13. Merelo-Guervós, J.J.: Agile (data) science: a (draft) manifesto. *CoRR abs/2104.12545* (2021). <https://arxiv.org/abs/2104.12545>
14. Merelo-Guervós, J.J.: What is a good doge? Analyzing the patrician social network of the republic of Venice (2022). <https://doi.org/10.48550/ARXIV.2209.07334>, <https://arxiv.org/abs/2209.07334>
15. Puga, D., Treffer, D.: International trade and institutional change: medieval Venice’s response to globalization. *Q. J. Econ.* **129**(2), 753–821 (2014). <https://doi.org/10.1093/qje/qju006>
16. Pullan, B.: ‘Three orders of inhabitants’: social hierarchies in the Republic of Venice. In: Denton, J. (ed.) *Orders And Hierarchies In Late Medieval And Renaissance Europe*. PFM, pp. 147–168. Macmillan Education UK, London (1999). https://doi.org/10.1007/978-1-349-27580-9_10
17. Rodrigues, F.A.: Network centrality: an introduction. In: Macau, E.E.N. (ed.) *A Mathematical Modeling Approach from Nonlinear Dynamics to Complex Systems*. NSC, vol. 22, pp. 177–196. Springer, Cham (2019). https://doi.org/10.1007/978-3-319-78512-7_10
18. Romano, D.: *The Limits of Kinship: Family Politics, Vendetta, and the State in Fifteenth-Century Venice*. Firenze University Press (2014)

19. Salehi-Abari, A., Boutilier, C.: Empathetic social choice on social networks. In: AAMAS, pp. 693–700 (2014)
20. Southworth, F.: Freight flow modeling in the United States. *Appl. Spat. Anal. Policy* **11**, 669–691 (2018)
21. Telek, Á.: Marrying the right one - evidence on social network effects in politics from the Venetian Republic (2017). https://editorialexpress.com/cgi-bin/conference/download.cgi?db_name=SAEe2017&paper_id=520
22. Zhao, Z., Zhu, X., Han, T., Chang, C.: The fake news propagate with more self-loop in online social networks. In: 2021 IEEE 2nd International Conference on Information Technology, Big Data and Artificial Intelligence (ICIBA), vol. 2, pp. 839–843 (2021). <https://doi.org/10.1109/ICIBA52610.2021.9688320>

Author Queries

Chapter 28

Query Refs.	Details Required	Author's response
AQ1	This is to inform you that corresponding author has been identified as per the information available in the Copyright form.	
AQ2	Please check and confirm if the authors Given and Family names have been correctly identified.	