A Social Network: Patterns of Co-authorship In The Field of Vigilance

California State University, Long Beach

Kelsey McCoy, Hanna Mac New, Xiaolu Bai, & Cindy Pham

April 28, 2020

Introduction

As the use of complex systems in society rises, it is important to understand the effects that inattention may have on outcomes. One example of complex systems that has increased greatly recently is the use of automated systems. As the use of automated systems increases in various sectors of the workplace, it is likely that human attention will decrease. That is, researchers predict that attention, specifically vigilance, will be negatively impacted as human operators are no longer required to monitor systems.

In the world of human factors, vigilance is defined as the ability to maintain concentrated attention over long periods of time (Neigel, et al., 2019; Hancock & Warm, 2003; Lyell, Magrabi & Coiera, 2018). Moreover, vigilance refers to the readiness of human operators to detect stimuli that are imperative to safety (Neigel, et al., 2019; Hancock & Warm, 2003). This paper will first explore the importance of vigilance, then will examine the network of researchers that are involved in research on vigilance. The aim of this paper will be to analyze the network of these researchers to better understand the connections and relationships that are present.

Importance of vigilance

The presence of vigilance is very fragile, as there are many factors that can potentially lead to its absence (Lyell et al., 2018). Once vigilance is lost, the operator is prone to make mistakes, and in complex situations, those mistakes often lead to injuries or fatalities. Such factors that impact vigilance include: time on task, fatigue, environmental characteristics, workload, and motivation (Hancock & Warm, 2003).

These factors are found to influence vigilance in a large variety of tasks. Therefore, research on this topic is relevant to a wide variety of sectors ranging from aviation to simple visual search tasks.

Network importance

As mentioned above, the topic of vigilance is of increasing importance as systems become more complex. Therefore, now, more than ever, it is critical to understand the network of researchers that are focused on this topic. Understanding such networks will allow researchers to expand their knowledge and work with a more diverse population.

Theoretical framework

The authors of this paper share a common interest in the field of Human Factors, with specific interest in vigilance and human performance. One reason for the interest in this subfield of human factors is that they have personal connections with professors at California State University, Long Beach (CSULB) who are or have researched this topic. After taking the Social Network Analysis course at CSULB, the authors thought it would be interesting to look at the social network of researchers on this topic. This would allow them to better understand who is conducting research on this topic outside of their direct connections.

The basis of this study was to see who is doing this research and what the relationships are between researchers. With their knowledge of the human factors field, the authors made the follow hypotheses about what their study would show:

1. The higher the author's impact factor is, the more closeness centrality they have.
2. Male authors have a higher degree centrality than female authors.
3. Male authors have a higher betweenness centrality than female authors.
4. Authors would more likely have co-authorship with people from their country.

Data and Methodology

There were four authors on this paper, and each author found roughly 28 articles. Thus, the resulting sample used in the research consisted of 115 digital articles. These digital articles were retrieved from the CSULB Library Database and Google Scholar. Articles were obtained based on their relevance to the topic of vigilance; to find such articles, key phrases were used to search for the papers including “vigilance”, “vigilance and automation”, and “vigilance in aviation”. In short, as long as the main focus of the publications referred to vigilance, it was included in the data. For the most part, the articles used were published within the past 10-15 years.

Organizing the data

With the finding of over 100 articles, it is critical to organize the data in a way that would lead to easy data analysis. Once all articles were obtained, categorical information was recorded in the excel network database. The first sheet of the excel database consisted of (1) article number, (2) article title, (3) the journal of publication, (4) that journal's average impact factor, and (5) all authors of the article. This sheet was used as the foundation for the construction of our social network analysis.

Following the recording of all 115 articles, a second excel sheet was created in the database. This sheet provided attributable information about individual authors; the recorded data consisted of (1) author number (1-309), (2) author name, (3) title of article(s), (4) the authors average impact factor, (5) gender, (6) school/organization affiliation, (7) state, (8) country, (9) highest degree, and (10) current position. Various pieces of information such as gender, country were manually searched using the Google database. Journal Citation Report (Incites) was used to look up the average impact factor.

A third, and final, excel sheet was created to show dyadic relationships between authors. This sheet consisted of: (1) author 1, (2) author 2, (3) up to seven columns for shared articles, and (4) average impact factor. This sheet was arguably the most important as it provided the relevant information to create our social network.

Once all the data was recorded into our excel database, we began to “clean” the data. This led to the exclusion of two articles as they both only had one author listed. The reason we had to exclude these articles is the fact that a relationship cannot exist between only one person. Once all of our data was recorded in our excel sheet, we identified the relevant attributes needed to test our hypotheses.

Identifying Relevant Attributes

In order to test the hypotheses, three attributes were kept for further analysis: average impact factor, gender, and country. Both gender and country were categorical variables, therefore we developed categorical codes for each. There were only two genders accounted for in our code, as all authors identified only as female or male; therefore, females were coded as “1” and males were coded as “2”. The second categorical variable, country, was based on where the author resided and was coded “1” through “24” (see table 1). The final variable of average impact factor was continuous and therefore not coded. Following the organizing of our variables, we used R statistical software (R Core Team, 2019) to run our data.

Table 1.

Countries and the corresponding codes used for analysis

Code	Country	Code	Country
1	USA	13	United Arab Emirates
2	Australia	14	Spain
3	Italy	15	Canada
4	France	16	UK
5	Germany	17	Chile
6	Taiwan	18	Algeria
7	Brazil	19	Sweden
8	Central Asia	20	Finland
9	Belgium	21	China
10	Netherlands	22	Tunisia
11	Korea	23	Israel
12	Singapore	24	Iran

The Statistical Analysis Process

The statistical analysis process was achieved through the use of R statistical software (R Core Team, 2019). Within R, a network analysis software package, igraph (Csárdi, 2020), was downloaded and utilized to compute degree centrality, betweenness centrality, and closeness centrality. Following these calculations, the values were then added to a fourth excel sheet within our database, to allow for organization. Another statistical package within the R software, HMISC (Harrell, 2020), was downloaded and used to find correlations between nodes and attributes. The last social network analysis software package within R that was downloaded, statenet (Handcock et.al, 2018), was used to analyze whether location was an influencing factor on co-authorship. To create a visualization of the network as a whole, the The function gplot was executed. A statistical overview of the attribute variables is provided in the following section.

Overview of Attributes Variables

First, the initial analysis showed that roughly 25% of all authors were female, leaving the remaining 75% of authors as male (see Table 2). At first glance, there appeared to be a great disparity of publications amongst gender in articles regarding vigilance. Next, it was found that nearly 51% of authors resided in the United States followed by Germany, Spain, and Tawin which constituted 22%, 14%, and 14% respectively (see Table 3). Without any true analysis, it seems as though the majority of articles related to vigilance are accounted for by authors in the United States. Lastly, our analysis found the mean of average impact factor to be *M* = 2.56. It is important to mention that our recorded average impact factors had a large range, from 0.041 to 5.812 (see Table 4). Many journals received an average impact factor of 2.649 (23%), followed by .8899 (7.2%) and 3.058 (6.6%).

Table 2.
Frequencies of Gender

Gender

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	1	76	24.9	24.9	24.9
	2	229	75.1	75.1	100.0
	Total	305	100.0	100.0	

Table 3.
Frequencies of Country

Country

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	1	155	50.8	50.8	50.8
	10	12	3.9	3.9	54.8
	11	3	1.0	1.0	55.7
	12	7	2.3	2.3	58.0
	13	5	1.6	1.6	59.7
	14	12	3.9	3.9	63.6
	15	4	1.3	1.3	64.9
	16	12	3.9	3.9	68.9
	17	2	.7	.7	69.5
	18	3	1.0	1.0	70.5
	19	4	1.3	1.3	71.8
	2	9	3.0	3.0	74.8
	20	2	.7	.7	75.4
	21	4	1.3	1.3	76.7
	22	2	.7	.7	77.4
	23	2	.7	.7	78.0
	3	7	2.3	2.3	80.3
	4	14	4.6	4.6	84.9
	5	22	7.2	7.2	92.1
	6	14	4.6	4.6	96.7
	7	2	.7	.7	97.4
	8	5	1.6	1.6	99.0
	9	3	1.0	1.0	100.0
	Total	305	100.0	100.0	

Table 4.
Descriptive Statistics of Average Impact Factor

	N	Range	Minimum	Maximum	Mean		Std. Deviation	Variance
	Statistic	Statistic	Statistic	Statistic	Statistic	Std. Error	Statistic	Statistic
average_impact	305	6	0	6	2.56	.070	1.231	1.515
Valid N (listwise)	305							

Analysis and Results

As previously mentioned, the network based data was analyzed using the igraph and statnet packages within the R statistical software. Initial analysis revealed that the network appeared to have multiple hubs, areas in which there are dense clusters of connections but they are isolated from each other (see Figure 1). In the center of the sociogram we see that there does appear to be a large cluster that is connected, by gatekeepers, to other large clusters. To better understand our initial sociogram, we will analyze the makeup of the network below.

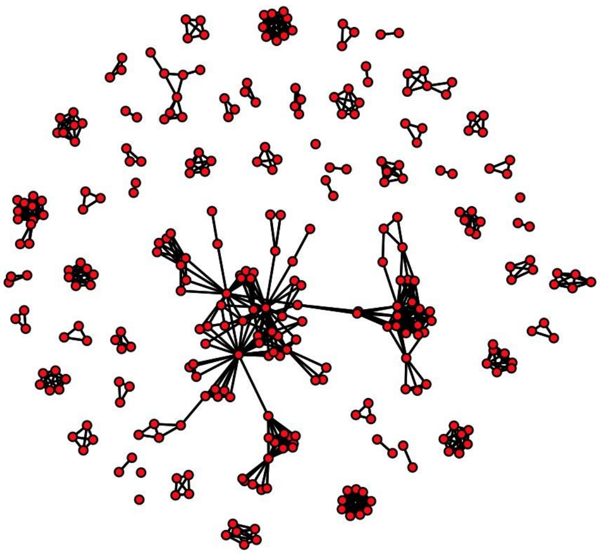

Figure 1.
Sociogram of the network.

Centralization

To understand the overall pattern of the network, the authors calculated three types of centrality of the network: degree, closeness, and betweenness. The degree centrality was used to measure the number of links an individual node has. For our particular dataset, nodes (authors) with a high degree centrality would be considered important or distinguished authors on the specific topic. Next, closeness centrality was used to measure how *close* each node was to each other node. The node with the highest closeness centrality would, theoretically, have the easister time passing on their ideas to the rest of the network. Lastly, we calculated betweenness centrality to measure how crucial one node is in the distribution of information. That is, nodes with higher betweenness centrality connect varying parts of the network together. Our analysis found that the node 13 had the highest centrality in all three types. The degree centrality of node 13 was 27, closeness centrality was 1.59e-05, and betweenness centrality was 2414.81. The results showed that node 13 was considered to be the most *central* node in the network. The data was further analyzed to test the four hypotheses stated in the earlier section.

Before examining the results of our specific hypothesis, it is important to provide the descriptive statistics of the centrality results. First, the degree centrality analysis showed that, on average, each node is linked to 4.8 other nodes. Second, the closeness centrality analysis showed that nodes were, for the most part, geographically far from one another, (M = 1.252e-05). Lastly, the betweenness centrality findings indicates that very few nodes are crucial in connecting various portions of the network (M = 51.07). In the following sections, we will analyze each of the hypotheses.

Hypothesis 1

To analyze our first hypothesis, which stated that "the higher the author's impact factor is, the more closeness centrality it will have", we used the Hmisc (Harrell, 2020) package within R. This hypothesis analyzed the relationship between average impact factor and closeness centrality, therefore a correlation was executed. The results of the Pearson correlation indicate that the authors' average impact factor is not significantly correlated with closeness centrality (r = -.08). This result suggests that, contrary to our prediction, authors with high impact factors are not more likely to have higher closeness centrality.

Hypothesis 2

Next, to analyze our second hypothesis, which stated that “male authors have a higher degree of centrality than female authors”, we, again, used the Hmisc (Harrell, 2020) package within R. This hypothesis analyzed the relationship between gender and degree centrality, again calling for the use of a correlation. The results showed that gender was not significantly correlated with degree centrality ($r = .05$). Therefore, gender was found to not have an effect on author’s degree centrality in this network.

Hypothesis 3

Next, to analyze our third hypothesis, which stated that “Male authors have a higher betweenness centrality than female authors”, we used the Hmisc (Harrell, 2020) package within R. This hypothesis analyzed the relationship between gender and betweenness centrality, once again calling for the implementation of a correlation. The results showed that gender was not, in fact, significantly correlated with betweenness centrality ($r = .03$). Like the previous two hypotheses, the result failed to prove that the gender of the author affects betweenness centrality.

Hypothesis 4

Lastly, to analyze our fourth hypothesis, which stated that “authors would more likely have co-authorship with people from their country”, we used the EGRM (Handcock et al., 2018) package within R. This hypothesis analyzed whether or not authors’ location was associated with co-authorship. Unlike the first three hypotheses, this hypothesis did not require the use of Pearson's correlation. Rather, the homophily index (2.416) and odds ratio (11.2) were calculated based on the country in which the author resided. The results indicated that authors who reside in the same country are 11 times more likely to co-author on a paper, which confirms our last hypothesis.

Table 5.
Betweenness centrality summary

Min	1st Quantile	Median	Mean	3rd Quantile	Max
0	0	0	51.07	0	2414.81

Table 6.
Closeness centrality summary

Min	1st Quantile	Median	Mean	3rd Quantile	Max
1.082e-05	1.089e-05	1.100e-05	1.252e-05	1.579e-05	1.585e-05

Table 7.
Degree centrality summary

Min	1st Quantile	Median	Mean	3rd Quantile	Max
1.0	2.0	4.0	4.8	6.0	27.0

Table 8.
Correlation table

	Average Impact Factor	**Gender**	**degree**	**betweenness**
Gender	0.06			
degree	0.02	0.05		
betweenness	-0.01	0.03	0.53***	
closeness_standardized	-0.08	0.09	0.36***	0.31***

Conclusion

To recap, the collected and evaluated data reflected articles of vigilance spanning topics such as aeronautics to task analysis. With such a broad topic, the collected data lacked sufficient power to indicate definitive results. Furthermore, with three of the four hypotheses resulting as insignificant, those results will be explored with a “what-if” lens. Below, we discuss our interpretation of each hypothesis findings, as well as offer insight for future studies.

Hypothesis 1

The first hypothesis, which examined author impact factor and closeness centrality resulted as insignificant ($r = -.08$). It was hypothesized that authors with articles in high impact factor journals would be more central in the network. Specifically, it was hypothesized that such authors would be connected to multiple other authors.

While conceptually this made sense, results indicated that impact factor did not lead to higher closeness centrality. With lack of significance, it is difficult to derive a concrete answer to this hypothesis. One idea, however, may be that authors with high impact factors are not, in fact, publishing often. Rather, these authors may spend prolonged time on one article, perfecting it, therefore having very few articles published in high impact factor journals.

A second idea for the insignificance of this hypothesis, may be that authors with high impact factors tend to only do research with other high impact factor authors. This would mean that high impact factor authors live in a small hub, not well connected to the rest of the network. Based on the statistical data that the majority of journals received an average impact factor of 2.649 (23%), followed by .8899 (7.2%) and 3.058 (6.6%), it can be inferred that authors with impact factors within the 2.5 range would have the highest closeness centrality.

Hypothesis 2

The second hypothesis, examining whether male authors have higher degree centrality than female authors, was also found to be insignificant, ($r = .05$). Based on the initial analysis, which showed an uneven distribution between the genders (with 75% males and 25% females), the impact of gender was brought into question. It was predicted that males would have more overall links in the network, considering the network make-up of gender.

This prediction made conceptual sense, but results indicated otherwise. In fact, the results indicated that neither men nor women had more ties (links) than the other. One possible explanation for this finding may be that male authors are likely to publish articles with few co-authors. That is, female authors have the same number of ties as men because they must work with more authors in order to get published. From our data, it can be seen that male authors did seem to publish with fewer co-authors than female authors, however this claim would need further investigation and analysis.

Hypothesis 3

The third hypothesis, consistent with the first two, found that gender and betweenness centrality were not significantly correlated ($r = .03$). We hypothesized that due to the uneven gender distribution, male authors would therefore have significantly higher betweenness centrality. This prediction was made in part due to our assumption that males would have higher overall centrality. However, we further assumed that male authors would be especially more likely to act as a bridge in the network when compared to female authors.

One possible explanation explored by the authors of this paper as to why this hypothesis is not true is related to the disadvantages faced by female publishing authors. In this case, even though there were far fewer female authors, they had to act as bridges as frequently as the males. It may be that female authors must act as bridges between other female authors and their male co-authors. This may even imply that male authors are more likely to work with multiple other male authors rather than multiple female authors.

Hypothesis 4

The final hypothesis investigated whether co-authorship was more likely to occur with individuals residing in the same country. Results indicated that this hypothesis was highly significant. These results were supported by the homophily index (2.416) and odds ratio (11.2) for the attribute of country. In order to better understand these findings, it is important to understand what both the homophily index and odds ratio are. First, homophily refers to the tendency in which people are more likely to create links and bond to those who share the same attributes (McPherson et. al, 2001). The index measures in-and-out group preferences, so it could be understood on which attributes are commonly shared between the nodes. The odds ratio of probability refers to strength of association between two events. Therefore, the odds ratio suggests that authors from the same country are roughly 11 times more likely to coauthor with one another.

The most basic explanation in support of this hypothesis is that proximity determines co-authorship. In regards to proximity, characteristics such as ability to speak the same language, understand cultural processes, and collaborate in the same time zone are assumed to increase the chances of co-authorship. The strong proximal coauthor relationships were consistent with those of Maisonneuve et al. (1952). In this study, students in a French boarding school had a tendency to mutually like one another as a result of past attendance of the same school and proximity of their desks. This further supports the idea that with close proximity, relationships are more likely to form. For the purpose of this study, relationships are from a professional standpoint.

With remote collaboration popularity increasing, the findings suggest that there may be a flaw in this type of communication. There was insufficient evidence to suggest a concrete reason as to why there is a lack of connection between authors of various countries. This could be due to social determinants, but there is also the possibility of a lack of usability in technology for communication (e.g. routers and devices). In fact, Figure 2 shows that the largest cluster of nodes is located in the United States. the author clusters of the data set with the largest being the United States. This can be explained by the data being obtained by Western based databases.

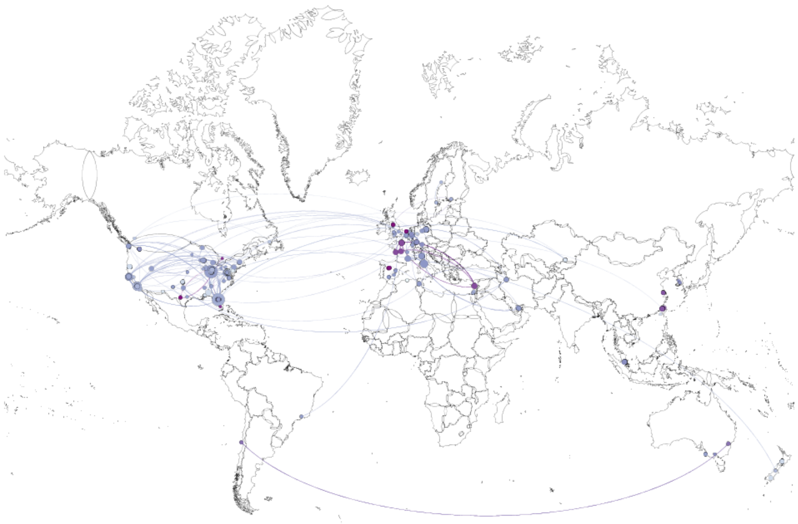

Figure 2.
The sociogram of the network based on author residence.

General Discussion

The findings of this study led to many assumptions based on the analysis of our data. While our first three hypotheses were found to be insignificant, we discussed that there may be underlying reasons as to why this occurred. The first hypothesis insignifiance may be explained by the idea that authors with high average impact factor tend to publish less frequently than authors with lower average impact factors. The second hypothesis led to the assumption that there may be unequal treatment between the genders. In fact, in an institutionalized system of societal norms, gender is an economic and sociocultural concept that disproportionately advantages men (Ridgeway & Correll, 2004). Moreover, Lundine et al. (2018) suggests that in the realm of academia, there is still evidence of the social disadvantage for women leading to the initial skepticism of male authors. Our final hypothesis was found to be significant, implying that authors residing in the same, or neighboring, countries are more likely to collaborate on articles. Future research on this network should aim to collect more articles in order to see if the same assumptions are indeed generalized.

References

Csardi, G., Nepusz, T. (2006). The igraph software package for complex network research. InterJournal Complex Systems 1695. http://igraph.sf.net

Hancock, P.A., & Warm, J.S. (2003). A dynamic model of stress and sustained attention. *Journal of Human Performance in Extreme Environments*, *7*(1).

Handcock, M., Hunter, D., Butts, C., Goodreau, S., Krivitsky, P., & Morris, M. (2018). ergm: Fit,

Simulate and Diagnose Exponential-Family Models for Networks. The Statnet Project (http://www.statnet.org). R package version 3.9.4, https://CRAN.R-project.org/package=ergm.

Harrell, F.E. (2020) Hmisc: Harrell Miscellaneous. R package version 4.4.0, https://CRAN.R-project.org/package=Hmisc

Lundine, J., Bourgeault, I.L., Clark, J., Heidari, S., Balabanova, D. (2018). The gendered system of academic publishing. *Lancet*. 391: 1754-1756.

Lyell, D., Magrabi, F., & Coiera, E. (2018). The effect of cognitive load and task complexity on automation bias in electronic prescribing. *Human Factors: The Journal of the Human Factors and Ergonomics Society*,*60*(7), 1008-1021.

Maisonneuve, J., Palmade, G., & Flourment, C.I. (1952) “Selective choices and propinquity ,” Sociometry, 15, 1-2, pp. 135-140.

McPherson, M., Smith-Lovin, L., & Cook, J. M. (2001). Birds of a Feather: Homophily in Social Networks. *Review of Sociology*, *27* 415-444.

Neigel, A.R., Dever, D.A., Claypoole, V.L., & Szalma, J.L. (2019). Task Engagement and the

Vigilance Decrement Revisited: Expanding Upon the Work of Joel S. Warm Using a Semantic Vigilance Paradigm. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *61*(3), 462–473. doi: 10.1177/0018720819835086

R Core Team (2019). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL http://www.R-project.org/.

Ridgeway, C.L. Correll, S.J. Unpacking the gender system a theoretical perspective on gender beliefs and social relations. Gend Soc. 2004; 18 : 510-531.