The Life of a Tie: Social Origins of Network Diversity

Patrick S. Park(✉) , Henry George Xu , and Kathleen M. Carley

Carnegie Mellon University, Pittsburgh, PA 15213, USA
{patrickp,hgx,kathleen.carley}@cs.cmu.edu

Abstract. This study examines the survival and evolution of 443K bidirectional mention ties on Twitter by merging datasets collected before 2015 and in the first few months of the COVID-19 pandemic (February to June, 2020). We hypothesize that strong pre-existing ties, marked by frequent communication and shared identities, endure and tolerate cognitive and stance differences over time. Our findings show that surviving ties are stronger than average pre-2015 ties but exhibit greater cognitive distance in COVID-19 discussions, suggesting that strong ties can tolerate different and even opposing opinions on contentious topics. This challenges traditional models of social influence and homophily, which predict increased cognitive similarity within strong ties. The findings imply the potential for old ties to function as network bridges, reducing political divides by connecting dissimilar social groups.

Keywords: strong ties · network diversity · tie survival · partisan bridging · covid-19

1 Introduction

Homophily is one of the most salient principles that give structure to social networks [11]. However, in almost any social context, social ties that diverge from this principle do exist - they tend to be weak network bridges that connect otherwise distant, dissimilar network communities. Viewed from the perspective of homophily, such non-homophilous ties are theoretically puzzling, yet the literature tends to explain away their presence with a functionalist logic, highlighting the information and control advantages that people potentially gain from bridging dissimilar groups [2,8]. This logic implies that individuals form and maintain non-homophilous ties in pursuit of such instrumental benefits [1,9]. However, recent studies show that bridging ties spanning long network distances may not necessarily rest on instrumental motivation - discussions between Twitter users who form a long-range tie tend to contain more words indicative of affective role relations (e.g., buddy, friend) and fewer words related to professional role relations (e.g., boss) [13]. The Chinese term, "guanxi", which generally denotes a particularistic relationship characterized by the warmth, trust, and obligation to another individual, can span dissimilar social communities, even in the

absence of structural support of common network neighbors [3]. How, then, do non-homophilous, structurally bridging ties come about?

Building on sociological insights highlighting the long-term cumulation of shared experience and history of a social tie for building trust [3,7], the current study explores the theory that non-homophilous bridging ties are likely the exceptionally strong ties that survive the long term, while their weaker adjacent ties decay with time [13]. A strong social tie may initially form in a tightly-knit group of people with similar attributes and in overlapping social contexts (i.e., high clustering and homophily with multiple shared foci) [6]. With the passage of time, group members tend to dissipate in social space to different schools, occupations, social classes, life styles, geographies, and political affiliations throughout the life course, leading the weaker ties that existed in the group to either naturally decay due to competing priorities or abruptly break when trust is breached. However, some of the strongest ties may endure these naturally compounding survival pressures of diverging life paths and experiences. The network neighbors who successfully sustain their strong ties over the long term may, perhaps unbeknownst to themselves, become the network bridges that structurally connect communities with different life styles, moral values, political beliefs, and group norms.

Our strategy for empirically testing this theory is to identify pairs of Twitter users that communicate over several years (@mention communication ties) and assess the changes in their relational strength, cognitive similarity, political disagreements, and the mutuality of communication. Specifically, among a sample of Twitter users who discussed COVID-19-specific topics with other users throughout the politically contentious months of the pandemic, we identify the @mention ties that had been observed previously in a large-scale Twitter data corpus collected before 2015. We analyze how the past relational and structural characteristics as well as the attribute similarities of these ties associate with those characteristics in the present and assess their capacity to maintain communication despite opposing stances on Covid vaccination, a hot-button issue in the U.S. that sparked fierce debates about individual freedom, public safety and health, and widening political divisions along pre-existing party lines. We hypothesize that communication partners who used to frequently communicate (i.e., strong ties), those who used to exhibit higher cognitive similarity, and those who shared similar identities (e.g., occupation, familial roles, political affiliation) and cultural interests (e.g., musical genre and sports) are more likely to maintain active bilateral communications over the long term. We further hypothesize that communication partners who likely initially developed the relationship based on similarity (i.e., homophily) will have developed tolerance for each other's differences in identity, cognition, and stance that tend to occur with the passage of time.

2   Data

We merge two temporally separated Twitter datasets to trace the survival and cognitive divergence of communication ties over time. The first dataset consists

of 26M U.S. Twitter users and their timelines of up to 3200 tweets, collected between 2013 and 2014 (the "pre-2015 data") [13]. The second dataset consists of a sample of tweets containing COVID-19 related keywords from February to June 2020 (the "Covid data"). We identify 443K bidirected mention ties (i.e., reciprocal mentions between two users) among 380K U.S. Twitter users in the pre-2015 data that also reappear in the more recent Covid data, either as unidirected or bidirected mention ties. These are the users who may have maintained their communications for five years or longer. While the pre-2015 data contain a more comprehensive set of tweets that a user created at the time, the Covid data covers only the Covid-specific tweets, limiting our observation of communications to COVID-19 subtopics. Nevertheless, it is possible to assess with these data how past relationships may affect communications in the long term on potentially sensitive topics during challenging times.

2.1 Variables

Tie Strength. We measure the strength of a tie by the frequency of mention tweets exchanged between two users, separately from both the pre-2015 data (past tie strength) and the Covid data (recent tie strength).

Tie Range. Following [13] and [8], we compute the second shortest path length of a tie from the pre-2015 data as a measure of the network structural distance that the tie bridges between otherwise disconnected network neighborhoods.

Cognitive Distance. We use a pre-trained 100-dimension GLoVe word embedding [14] to map each user's cognitive location in high-dimensional vector space by aggregating the word vectors from their respective tweets. We subsequently measure the cognitive distance between users by computing the Euclidean distance of their respective aggregate word vectors. This cognitive distance reflects a broad range of differences in knowledge, interests, and beliefs, and can also be viewed as a measure of (the inverse of) cognitive homophily. By computing this cognitive distance of a mention tie separately from the two datasets, we gauge the relative growth or reduction in cognitive distance over time.

Stance Distance. The cognitive distance between two users based on the GLoVe word vectors may simply reflect *different* interests and not necessarily *opposing* stances on contested issues. Therefore, we measure the users' stance on COVID vaccination and compute the difference in stances between two users as an indication of opposing positions that a tie may be tolerating. Specifically, from the Covid dataset, we use a pre-trained COVID-Twitter-BERT model [12] for tweet-level vaccination stance detection. The last layer hidden state computed from this model is used as an abstract vector representation of the vaccination-related stance expressed in a tweet. We aggregate these tweet-level stance representations at the individual level and use it as a summary vector representation

of the individual's overall stance on vaccination. Although this approach does not allow straight forward qualitative interpretation of a user's stance (e.g., pro- vs. anti-vaccination), it is a richer representation that renders a continuous, granular measure of stance opposition between two users. Although the hidden layer representation does not provide a readily human-readable indicator, its correspondence to the intuitive categorical labels are straightforward as shown in Fig. 1, which demonstrates the clear separation and groupings in both PCA and t-SNE plots.
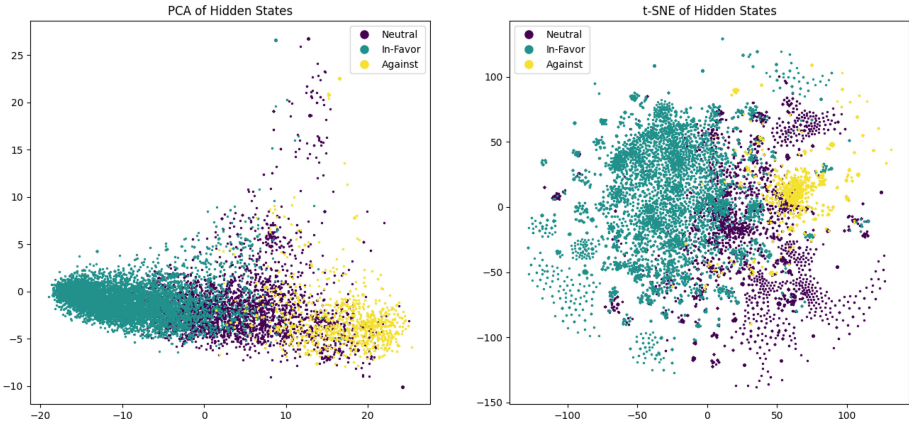

Fig. 1. PCA and t-SNE visualizations of hidden states. Colors indicate different stance labels on Covid vaccination, with each cluster representing a distinct stance group

Identity Overlap. Apart from the cognitive (dis)similarity of two users, we explore how the similarity in users' salient past identities (i.e., identity homophily) and cultural interests correlate with the survival of their commu- nication ties. Specifically, we construct a prototype identity lexicon[1] for the categories of occupation (e.g., "reporter" at Boston Globe), familial roles (e.g., proud "father"), political affiliation (life-long "democrat"), interest in sports (e.g., football), and cultural consumption (e.g., "punkrock", "historybuff") and code the occurrence of these terms in the user profile descriptions. Based on this identity and interest codings of each user's profile description, we construct category-specific dichotomous variables that indicate whether two users who share a tie both listed the same identity terms or not in their pre-2015 profile descriptions.

---

[1] Lexicon and keyword occurrences available: https://doi.org/10.5281/zenodo. 11430935.

2.2   Analytical Strategy

In the first set of analyses, we describe the characteristics of the pre-2015 bidi-rected mention ties that reappear in the Covid dataset either as unidirected or bidirected mention ties, in terms of their tie strengths, cognitive distances, and vaccination stance distances. Based on these quantitative descriptions, we assess the likelihood of the pre-2015 bidirected tie to remain in bidirected communica-tions (i.e., tie survival), relative to one-way (unidirected) communications in the Covid data. Here, we use logistic regression to evaluate the odds of a pre-2015 tie showing up as bidirected in the Covid dataset with other tie-level covariates, including pre-2015 tie strength and shared identities on each identity category.

3   Results

Consistent with our assumption that strong ties are more likely to survive, the ties observed in the pre-2015 data that reappear in the Covid data tend to be relationally stronger than the average pre-2015 tie. As shown in Fig. 2, these "old ties" exchanged approximately 40% more mention tweets than the average tie in the pre-2015 data (red dashed line), irrespective of the network distance (i.e., tie range) they spanned in the past.

Figure 3 plots the mean cognitive distance separately measured from the two datasets. Since the Covid dataset was constructed from tweets about a narrower issue than the issue-agnostic pre-2015 tweets, the former exhibits a shorter cogni-tive distance (orange) than the distance in the general, topic-agnostic tweets from the pre-2015 data (blue) as one would expect. Furthermore, while the pre-2015 cognitive distances tend to increase with tie range as expected, more noteworthy is the same pattern emerging in the Covid cognitive distances. That is, the tie range from years prior continues to correlate with the cognitive distance around Covid.

Figure 4 plots the relationship between cognitive distance (y-axis) and tie strength (x-axis), crossed by the time of measurement (pre-2015 vs. Covid). Panels A and D (main diagonal) exhibit associations between cognitive distance and tie strength that are consistent with homophily - when measured in the same time periods, the users who engage in more frequent conversations also tend to exhibit greater cognitive similarity at that time. Surprisingly, however, the ties that are cognitively distant in their Covid-related discussions used to engage in more frequent communications pre-2015 (panel B), suggesting that the relationally strong ties that survive over time might develop the trust and toler-ance to discuss their differing conceptions on such contentious issues as Covid. Conversely, the old ties that engage in more frequent Covid-related conversa-tions used to exhibit greater cognitive distance pre-2015 (panel C), a result that is not readily explainable by homophily.

Granted, cognitive distance does not necessarily indicate that two users hold opposing stances on contentious issues. For example, one user might tweet mostly about the psychological isolation of the lockdown while the other might tweet about the supply shortages triggered by the lockdown. Therefore, in Fig. 5,
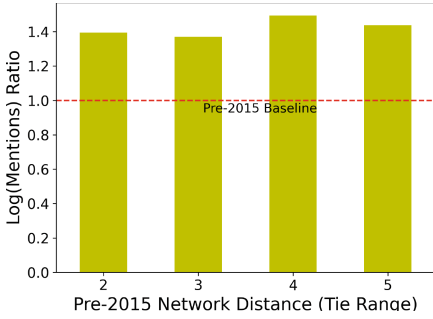
Fig. 2. Old ties observed in COVID-19 tweets (survived ties) tend to be stronger than the average old tie (red horizontal line). Tie range is the second shortest path length of a tie, which measures the network distance that the tie spans (Color figure online)
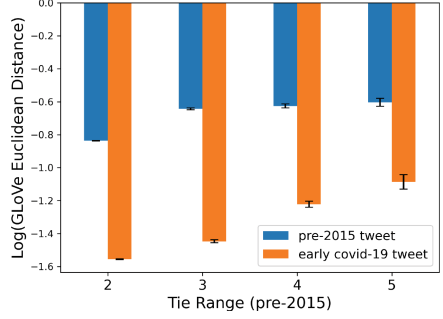
Fig. 3. Cognitive distance of old ties. Compared to the cognitive distance measured in their dyadically exchanged tweets pre-2015 (blue bars), the cognitive distance in the COVID-19 tweets they exchanged (orange bar) tends to be shorter (Color figure online)
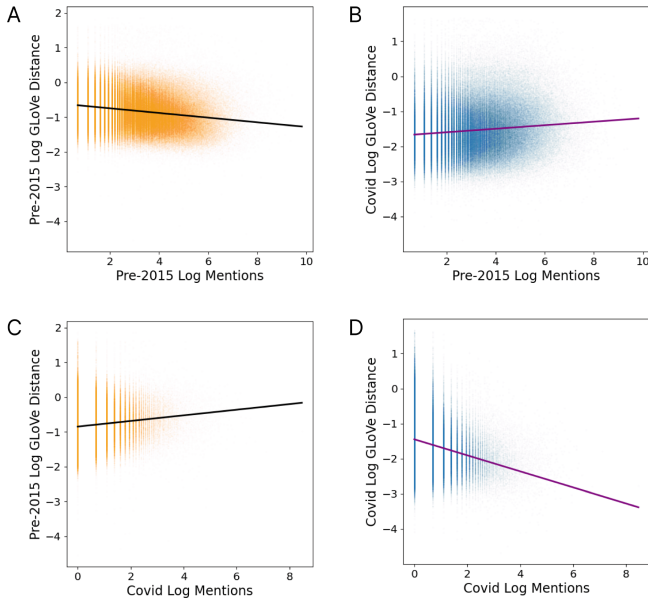

Fig. 4. Cognitive distance in conversations between old ties (pre-2015 and COVID data) and their relational strength measured by mention frequency (pre-2015 and COVID data). The strong ties in pre-2015 tend to show greater cognitive distances in their COVID conversations (panel B)

we directly measure stance opposition on the issue of Covid vaccination and explore its association with the a tie's relational strengths, measured from the two datasets. Similar to the counter-intuitive increase in Covid-related cognitive distance with the increase in pre-2015 mention frequency (panel B of Fig. 4), vaccination stance distances are not shorter for the strong ties in pre-2015 (left panel of Fig. 5) as one might expect to observe with the tie strength in the Covid data (right panel). In short, people who used to have a stronger tie in the past do not appear to take more similar stances on Covid vaccination. In fact, we find that these previously strong ties can tolerate broader stance differences, albeit at lighter levels of engagement as shown in Fig. 6 (orange line).
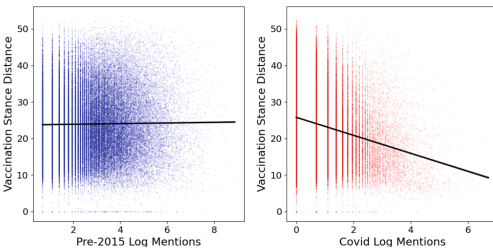

Fig. 5. Stance distance on COVID-19 vaccination does not vary by the Old ties' pre-2015 tie strength (left panel). On the other hand, stance distance decreases with frequent mentions in their COVID-specific tweets (right panel)
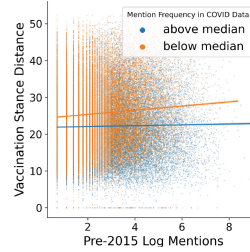
Fig. 6. Relationally stronger old ties tend to tolerate wider stance distances on COVID vaccination, more so when discussed less frequently (orange) (Color figure online)

We find further evidence for the critical importance of tie strength in maintaining a lively interpersonal tie in the long-run. Table 2 reports logistic regression for the likelihood that a pre-2015 tie (all bidirected) continues to engage in bidirected communications (i.e., stricter definition of a survived tie) vs. unidirected communications in the Covid data. We use the pre-2015 log mention frequency and the matches in the users' salient identities and interests across five different categories as predictors. To address the severe class imbalance in the distributions of these five dichotomous variables as shown in Table 1, we further ran a logistic regression model with SMOTE (Synthetic Minority Over-sampling Technique) [4], but the direction and statistical significance of the results were qualitatively similar to the simpler model reported in Table 2. Net of identity and interest overlap, a unit increase in logged mention frequency is associated with a 39% ($\exp(0.328) = 1.39$) increase in the odds of a bidirected tie. On the other hand, the similarities in identity and interest between two old network neighbors do not consistently predict an increase or decrease in the odds of a bidirected mention tie in the Covid data. Specifically, the odds for users who both listed the same family roles (e.g., "father") were higher by 114% than those whose family roles did not match, either due to mismatches in those terms

Table 1. Tie Level Identity Matches

Shared Identity/Interest Category	Percentage
Same Occupation	5.12
Same Family Role	0.31
Same Political Orientation	0.53
Same Cultural Interests	1.91
Same Sports Interests	0.92

Table 2. Logistic Regression on Bidirected vs. Unidirected Ties in the Covid Data

Variable	B	S.E.	
Log Mention Frequency (pre-2015)	0.33	0.02	**
Same Occupation	−0.04	0.09	
Same Family Role	0.76	0.06	**
Same Political Orientation	−0.10	0.04	*
Same Cultural Interest	−0.41	0.06	**
Same Sports Interest	−3.02	0.01	**

$^{*} < 0.01$, $^{**} < 0.001$ N $= 425597$ $R^2 = 0.017$

or to their absence in the descriptions. Matches in occupational (e.g., "journalist") and political (e.g., "progressive") identities, in contrast, were both far more weakly associated with bidirected ties, each rather decreasing the odds by 4.05% and 9.55%, respectively. The match in cultural interests exhibited even stronger negative associations. For example, two users with the same cultural interest were 33.68% less likely while fans of a same sport were 95.10% less likely to maintain a bidirected tie. In short, the abstract notion that similarity begets friendship does not apply consistently across different social, political, and cultural dimensions on which people find commonality and maintain relationships in the longer term.

4   Discussions

As expected, the ties that appear to survive the long term are relationally strong. The old ties in the Covid data tend to be the strong ties and the bidirected ties in the covid data tend to be stronger than their unidirected counterparts. However, these survived ties exhibit qualitative characteristics that are theoretically puzzling in light of the central position homophily assumes in our understanding of network dynamics. Our analyses show that strong pre-2015 ties exhibit (a) greater cognitive distance in their Covid-related conversations and (b) no systematic difference in their vaccination stance distance. These findings are theoretically puzzling in light of prominent opinion dynamics models that emphasize the positive feedback loop between social influence and homophily [10] - actors with a strong tie exert stronger social influence on each other to become more similar and, in turn, greater similarity subsequently strengthens the social tie even further.

Our tie-level results at longer time scales also call into question the opposite side of this theoretical feedback loop between social influence and homophily. That is, just as strong ties in the past exhibit greater cognitive distance on Covid-related issues, we also find that ties with shared identities and/or interests in the past do not necessarily associate with a higher probability of mutually active (i.e., bidirected) communications regarding Covid. Specifically, except for when users described themselves with same familial roles, users with matched

political and occupational identities in the past tend to have lower, albeit with weaker effect, odds of bidirected communications about Covid. Furthermore, the odds of mutual communications were unambiguously lower for ties with common cultural and sports interests in the past. One possibility is the potential cognitive dissonance that old ties might experience when their past shared identities and interests are overshadowed by the new identities and interests developed separately over time. If these disjoint identities and interests align along the divisions in opinions and stances about Covid, but crosscut their prior shared identities and interests [5], it is conceivable that these users avoid engaging each other in uncomfortable disagreements about Covid. However, the longer-lasting, "stickier" social identities, such as family role identities (e.g., identity as a father generally lasts longer than one's occupational identity), might not cause such cognitive dissonance as such social ties can experience similar life course events more or less concurrently *and* develop similar attitudes and opinions from those similar experiences (e.g., two mothers may share similar opinions about vaccination, mask mandates, and school reopenings based on the same unprecedented experience of starting to send their children to middle school during the Covid lockdown). Perhaps this is why the familial identity is the only positive coefficient in the logistic regression.

From a practical standpoint, our results suggest the possibility of these old strong ties functioning as network bridges that span social groups holding different cognitions and/or opposing views on contentious topics. Although further research is needed to ascertain the real informational and affective bridging effects of old ties, if their general efficacy for reducing polarization is supported by accumulated evidence, social media platforms may be able effect simple changes that support the long-term communications between platform users.

Although this study makes an important methodological contribution to behaviorally studying social and communication ties over long time frames, it also carries important limitations. First, our observation of long-surviving ties takes only two snapshots, thereby leaving a long temporal blindspot in between. Therefore, it is possible that the ties we observe in the Covid data are not sustained throughout the years, but rekindled due to the highly unusual circumstances created by the pandemic (e.g., increased remote communication due to the lockdowns). Related, the observed ties are highly specific to COVID-19 issues. Therefore, even if some ties in the data had sustained long-term communication on other topics, they are unobservable from our data. Although this severely limits the generalizability of our findings, our preliminary replications on a dataset not related to Covid (i.e., the 2021 Canadian Federal Election) show agreement with the results reported in the current study. Future extensions would need to address these issues by applying a similar study design to Twitter datasets collected on other topics.

References

1. Burt, R.S., Jannotta, J.E., Mahoney, J.T.: Personality correlates of structural holes. Soc. Netw. **20**(1), 63–87 (1998). https://doi.org/10.1016/S0378-8733(97)00005-1, https://www.sciencedirect.com/science/article/pii/S0378873397000051
2. Burt, R.S., Kilduff, M., Tasselli, S.: Social network analysis: foundations and frontiers on advantage. Annu. Rev. Psychol. **64**(1), 527–547 (2013). https://doi.org/10.1146/annurev-psych-113011-143828, https://doi.org/10.1146/annurev-psych-113011-143828. pMID: 23282056
3. Burt, R.S., Opper, S.: Guanxi and structural holes: strong bridges from relational embedding. Am. J. Sociol. (ja), null (2024). https://doi.org/10.1086/730630, https://doi.org/10.1086/730630
4. Chawla, N.V., Bowyer, K.W., Hall, L.O., Kegelmeyer, W.P.: SMOTE: synthetic minority over-sampling technique. J. Artif. Intell. Res. **16**, 321–357 (2002). https://doi.org/10.1613/jair.953
5. DellaPosta, D., Shi, Y., Macy, M.: Why do liberals drink lattes? Am. J. Sociol. **120**(5), 1473–1511 (2015). https://doi.org/10.1086/681254
6. Feld, S.L.: The focused organization of social ties. Am. J. Sociol.**86**, 1015–1035 (1981), http://www.jstor.org/stable/2778746
7. Granovetter, M.: Problems of explanation in economic sociology. Netw. Organ.: Struct. Form Action 25–56 (1992). https://cir.nii.ac.jp/crid/1570291225791515520
8. Granovetter, M.S.: The strength of weak ties. Am. J. Sociol. **78**, 1360–1380 (1973). http://www.jstor.org/stable/2776392
9. Lyu, D., Yuan, Y., Wang, L., Wang, X., Pentland, A.: Investigating and modeling the dynamics of long ties. Commun. Phys. **5**, 87 (2022). https://doi.org/10.1038/s42005-022-00863-w
10. Macy, M., Kitts, J., Flache, A., Benard, S.: Polarization in dynamic networks: a hopfield model of emergent structure, pp. 162–173 (2003)
11. McPherson, M., Smith-Lovin, L., Cook, J.M.: Birds of a feather: homophily in social networks. Annu. Rev. Sociol. **27**, 415–444 (2001). https://doi.org/10.1146/annurev.soc.27.1.415, https://www.annualreviews.org/content/journals/10.1146/annurev.soc.27.1.415
12. Müller, M., SalathÃ, M., Kummervold, P.E.: Covid-twitter-BERT: a natural language processing model to analyse Covid-19 content on twitter. Front. Artif. Intell. **6** (2023).https://doi.org/10.3389/frai.2023.1023281, https://www.frontiersin.org/articles/10.3389/frai.2023.1023281
13. Park, P.S., Blumenstock, J.E., Macy, M.W.: The strength of long-range ties in population-scale social networks. Science **362**, 1410–1413 (2018). https://doi.org/10.1126/science.aau9735, http://science.sciencemag.org/content/362/6421/1410.abstract
14. Pennington, J., Socher, R., Manning, C.D.: GloVe: global vectors for word representation, pp. 1532–1543 (2014). http://www.aclweb.org/anthology/D14-1162