PNAS nexus

PNAS Nexus, 2024, **3**, pgae549
https://doi.org/10.1093/pnasnexus/pgae549
Advance access publication 10 December 2024
Research Report

Complements and competitors: Examining technological co-diffusion and relatedness on a collaborative coding platform

Antonio D. Sirianni[a,b,*], Jonathan H. Morgan[c,d], Nikolas Zöller[d,e], Kimberly B. Rogers[a] and Tobias Schröder[d]

[a]Department of Sociology, Dartmouth College, 20 N Main St, Hanover, NH 03755, USA
[b]McCourt School of Public Policy, Georgetown University, 125 E St NW, Washington, DC 20001, USA
[c]Department of Sociology, Duke University, 417 Chapel Drive, Durham, NC 27708, USA
[d]Institute for Applied Research Urban Future, Potsdam University of Applied Sciences, Kiepenheuerallee 5, Potsdam 14469, Germany
[e]Department of Adaptive Rationality, Max Planck Institute for Human Development, Lentzeallee 94, Berlin 14195, Germany
*To whom correspondence should be addressed: Email: as5064@georgetown.edu
Edited By Attila Szolnoki

Abstract

Diffusive and contagious processes spread in the context of one another in connected populations. Diffusions may be more likely to pass through portions of a network where compatible diffusions are already present. We examine this by incorporating the concept of "relatedness" from the economic complexity literature into a network co-diffusion model. Building on the "product space" concept used in this work, we consider technologies themselves as nodes in "product networks," where edges define relationships between products. Specifically, coding languages on GitHub, an online platform for collaborative coding, are considered. From rates of language co-occurrence in coding projects, we calculate rates of functional cohesion and functional equivalence for each pair of languages. From rates of how individuals adopt and abandon coding languages over time, we calculate measures of complementary diffusion and substitutive diffusion for each pair of languages relative to one another. Consistent with the principle of relatedness, network regression techniques (MR-QAP) reveal strong evidence that functional cohesion positively predicts complementary diffusion. We also find limited evidence that functional equivalence predicts substitutive (competitive) diffusion. Results support the broader finding that functional dependencies between diffusive processes will dictate how said processes spread relative to one another across a population of potential adopters.

Keywords: networks, diffusion, innovation, relatedness, Computational Social Science

Significance Statement

This article assesses how a set of related technologies co-diffuse with one another by looking at how the diffusion of users between innovations relates to the compatibility of innovations. Building on both prior studies of network diffusion and research on "relatedness" from the economic complexity literature, we use publicly available GitHub data from millions of users and coding projects to examine how the co-diffusion of coding languages relates to how frequently they appear alongside one another in the same project, or if they tend to serve as substitutes for one another.

Introduction

How do technologies spread with and against one another? Most prior work considers whether individuals adopt an individual technology as it spreads across a network. Moreover, this tends to focus primarily on new entrants, or innovations, within a broader set of interconnected technological products. Here, we consider how new technologies, and products more generally, may grow with and against each other in a network of their own. This builds on previous research studying competitive and overlapping contagious processes and by incorporating measures of how diffusing products functionally relate to one another, similar to the notion of "product space" from the literature on economic geography and complexity. The degree to which different technological products are used in conjunction or in place of one another affects how languages share or transfer users.

Network models of innovation and diffusion

Research on innovation diffusion is essentially research on how products diffuse with respect to the beginning of their life cycle and is often the focus of network-based analysis. The prototypical

Competing Interest: The authors declare no competing interests.
Received: June 19, 2024. **Accepted:** November 19, 2024

diffusion process involves the gradual spread of a possibly contagious phenomenon such as an idea, technology, or disease, through a certain proportion of a population (1). When someone catches a virus or accepts an idea, they become a vehicle of transmission, accelerating growth until the decaying base of potential recruits decelerates growth. The diffusion curve generated by this process provides a good null model, but the innovation trajectory is largely influenced by the characteristics of innovations, the characteristics of innovators, and the environmental context (2). More sophisticated models will incorporate the mechanisms that are unique to each diffusion. For example, a disease can simply jump from one person to another as a result of physical proximity, whereas a belief may require social interactions with multiple individuals and a conscious decision by the recipient to spread. Contagious processes that require multiple points of contact are referred to as "complex" (3, 4).

An additional layer of complexity is introduced when the interaction of diffusive processes is considered. As has been shown with the transmission of infectious diseases, the spread of one contagion may facilitate the spread of another (5–7), and the spread of information and awareness may inhibit their diffusion, albeit indirectly via altering the behaviors of individuals. (8, 9). Similar contagion dynamics may affect products and technological innovations, one example being the co-diffusion of UPC codes and Scanners (10). Examining the world of product diffusion in general invites the consideration of competing products and innovations, one popular case being the competition between VHS and Betamax in the 1980s (11–13). Competitive co-diffusion has also been examined using a simulation-based approach (14).

Some examinations of interacting diffusive processes have scaled up to look at how larger sets of contagions can interact with each other. Simulation-based analyses of multiple competing contagions have examined how the limited cognitive space and network structure of individual actors can inherently cause heterogeneous and polarized levels of success within a set of competing contagions (15, 16), or considered how different contagion types (i.e. simple vs. complex) can be temporally intertwined in the same diffusive process and become codependent (17, 18). Empirical work has also shown that contagions in online spaces can become intertwined (19, 20). (For a more comprehensive review of co-evolving diffusive processes, see (21).) While such cases are illuminating, they do not account for the complex relationships that may characterize a larger set of diffusing technologies in which the focal innovations are embedded. Furthermore, large sets of technologies may have substitutive relationships (VHS and Betamax) or complementary relationships (UPC codes and scanners) with one another, providing an opportunity to observe larger networks of interacting diffusive processes.

Product space and relatedness

Considering a larger set of phenomena, we can envision a network of connected contagions at a level of abstraction beyond the network of connected individuals. The literature on economic geography and complexity provides a promising direction for how a set of interacting and intersecting products can be considered in the context of one another. The notion of a "product space" was introduced to describe which products require similar infrastructures, and how nations or regions that are advantaged or disadvantaged by their position in this space could grow and transition their mix of economic exports (22, 23). The position of a nation's industrial capacity in product space can set it on a certain trajectory, where countries with complex economies that occupy high-density areas in the core of product space, as well as a diversity of positions, are well positioned for future growth (24, 25).

The general principle that an individual, firm, region, nation, or other entity will embrace a new productive activity with a probability that is proportional to activities they already participate in is referred to as "relatedness" (26, 27). This finding has been validated in a variety of contexts. In a context of technological innovation and production, technologies that are related to each other through patents are likely to emerge in the same regions (28, 29), the product space of a region will affect its ability to develop patents (30), and new product types are likely to emerge in areas close to regional centers of industrial structure (31). Similarly, studies of labor and geography have revealed that occupations that are linked through shared skills are likely to emerge in the same job markets (32, 33). This principle also applies to the world of knowledge production; branches of scholarship that are related through overlapping scholars are likely to emerge in the same universities (34).

Certain parts of product space may be beneficial to the products themselves, not just the producing actors that work in those spaces. For example, new scientific ideas that are embedded with a set of established concepts tend to diffuse more widely (35). Similarly, certain ideas, products, or practices may be more or less compatible with one another, and an individual's adoption of one may make them more or less susceptible to others. Focusing specifically on the context of innovation and the co-diffusion of new technologies, certain products may be explicitly designed to build on (or undermine) the success of other products and spread across their existing user bases. Products may be designed to target another technology by either trying to replace it or by trying to complement it and ride on the coattails of its success. Alternatively, a new innovation may unintentionally exploit or replace existing technologies. Whether functional relationships between technologies are intentional or incidental, the presence of one technology can change the costs or benefits of adopting other technologies. Different positions in a constantly evolving product space may provide opportunities for growth, vulnerability to decline, or both.

Collaborative coding as an empirical case

While a primary aim of the product space literature is to explain which parts of the space can accelerate or hinder the growth of national or regional economies, our concern is with the diffusion and remission of the products themselves. The vast number of computer programming languages that have emerged over the years offers a compelling case study for a set of competing and co-operating technological products. Although spoken languages have been considered in previous work on diffusion (36), the frequency and speed with which programming languages can be adopted offers a far more dynamic and data-rich context for examination. Integrating prior work that has considered the case of computer language co-development as a site of competitive innovations (37, 38) and examined product spaces of software development (39), our aim is to demonstrate that the diffusion of technological products is related to how they are functionally positioned in a broader ecology of adjacent technologies.

An extensive empirical dataset available from the GitHub public repository enables an examination of diffusing products on a grand scale. In contrast to other diffusion research, we do not use this data to examine how innovations spread relative to a network of interconnected individual users (although many of them certainly are connected to one another in one way or another).

Instead, we analyze how these languages build on each other and detract from each other and constitute a broader network of technologies. We test the overall network of languages for the presence of two types of diffusion: substitutive (or competitive) diffusion where one product (language) tends to replace and impede the spread of another, and complementary diffusion where one product spreads disproportionately in the presence of another language.

The aforementioned studies of relatedness and product space often focus on relatedness as it pertains to the success of the producer. The unit of observation is a distinct social or economic unit, the location in product space is an independent variable, and the dependent variable is an economy, firm, or sector's adoption of a product. Instead, we use interrelated networks of products themselves (an expression of product space) to explore how potential users are shared and stolen between different products, making the products themselves (and the connections between them) the primary units of observation.

Our measures of diffusion are dyadic, and depend on how users of a language at time t adapt another language at time $t + 1$. These rates of diffusion between languages are modeled on the basis of how pairs of languages are used together at time t. Drawing on fundamental work in social network analysis that connects individuals based on shared group memberships (40), we use bipartite relationships between languages and projects to determine how languages are functionally connected to one another. From this network of connections, we draw on earlier work emphasizing the difference between the structural equivalence of nodes (where nodes have similar sets of connections) and cohesion (where nodes are connected to one another) in social networks (41, 42), to create and quantify concepts of functional equivalence and cohesion. Functionally equivalent languages frequently co-appear with the same sets of other languages, implying they may be substitutes, whereas functionally cohesive languages frequently appear with one another, suggesting that the use of one enables or enhances the use of another. Similar complementarity and substitutability frameworks have been used to think about market competition strategies (43) and to improve recommendation approaches in large product spaces (44, 45). Measuring these concepts of co-diffusion and co-functionality within a large public and collaborative online working space provides a rare portrait of how large sets of interrelated technologies percolate with one another in a broader community.

Data

The data we use come from the GHtorrent database (originally available at GHtorrent.org), which was originally retrieved from the GitHub API. GitHub is a social coding platform where groups and individuals collaboratively edit shared coding projects, although many projects feature only one individual programmer. These data have been used extensively to study processes of group collaboration, emergent hierarchies, and meaning-making (46, 47). GitHub data have also helped illuminate how gendered inequalities emerge in technology work groups (48, 49), and the role of gender diversity in group success (50). This study takes a wider lens and focuses on the broader patterns of how programming languages spread relative to one another as a whole. We focus on a subset of the data from the beginning of 2010 to the end of 2018. For each project that is entered in the GitHub dataset, an initial user, date, and set of languages are available. This allows us to construct one data set that summarizes how languages co-occur with one another across projects, and another data set which contains the history of yearly language use for individual GitHub users.

The co-occurrence of languages within projects is independently assessed for each year and indicates which languages are frequently used with one another (functional cohesion) and which languages may be interchangeable with one another (functional equivalence). The language usage patterns of individuals in aggregate provide a measure of how languages diffuse with (complementary diffusion) and against (substitutive or competitive diffusion) one another in the broader GitHub community. These diffusion measurements are dynamic and require data from two adjacent time periods, each one year in length. A time window length of a year is short enough to capture temporal changes in mechanisms (and also allows the estimation of several models as a robustness check), but long enough to allow users of a language in any time window t sufficient time to use that language again in time $t + 1$. It also eliminates any potential issue of seasonal variation in coding behavior and activity. Specifically, the patterns of how languages are used together on projects during one time period, t, is used to predict how language users will begin using new languages between time period t and time period $t + 1$. A quantitative summary of the total number of languages, projects, and users in each of the eight overlapping 2-year windows is provided in Table 1. A visual summary of the proposed types of co-functionality and diffusion, and the hypothesized relationships between them, is provided in Fig. 1.

Hypotheses

Given the data and context at hand, we expect network co-functionality and diffusion to relate to one another in the following ways:

- H1: Functional cohesion between two languages will predict higher rates of complementary diffusion. The use of one language will lead to the use of the other in a later time period. (This hypothesis corresponds most directly to the principle of relatedness in the data.)
- H2: Functional equivalence between two languages will positively predict competitive substitutive diffusion. The interchangeability of languages will cause their users to be more likely to switch from one to the other.
- H3: Functional cohesion will negatively predict substitutive (competitive) diffusion. If two languages are frequently used with one another they will be unlikely to undermine each other's user base. (This hypothesis is a converse implication of relatedness.)
- H4: Functional equivalence will positively (H4a) or negatively (H4b) predict complementary diffusion. We do not have a single prediction on whether functional equivalence will be positively or negatively associated with complementary diffusion. Functional equivalence may indicate that two languages are used for the same purpose, thus domain experts in said purpose may add more languages to their repertoire. On the other hand, if a user knows how to complete a task with one language, they may gain little benefit from learning another. (This relationship is considered using two-tailed criteria for statistical significance.)

Methods

Determining the relationship between diffusion and functionality requires us to consider how languages are both shared by projects, which shows the extent to which they are used together, and by individuals, which shows how languages spread between and

Table 1. Summary of data analyzed in the GitHub dataset—for each overlapping 2-year window, summaries of total (multilanguage) projects and average languages per (multilanguage) project are shown for year t, the users active in both years t and $t+1$, and the number of languages that are included in the final network analysis across both years.

Year	Year	Multiple language	Languages per	Users in	Languages in final network		
t	t+1	projects in t	project in t	t & t+1	(0.5%)	(1%)	(2%)
2010	2011	45,440	3.45	53,822	23	20	13
2011	2012	131,849	3.56	127,130	24	20	12
2012	2013	373,865	3.44	336,115	25	17	14
2013	2014	886,170	3.46	603,260	30	18	15
2014	2015	1,620,247	3.49	928,088	28	19	16
2015	2016	4,781,850	3.93	1,478,089	29	20	15
2016	2017	2,805,822	3.80	2,217,855	25	17	14
2017	2018	5,931,636	4.04	2,937,140	26	16	14

Languages with fewer than 0.5, 1, or 2% of users in year t are excluded from final network models of diffusion for that period. (Separate models are estimated with each of these three cutoffs as a robustness check.)

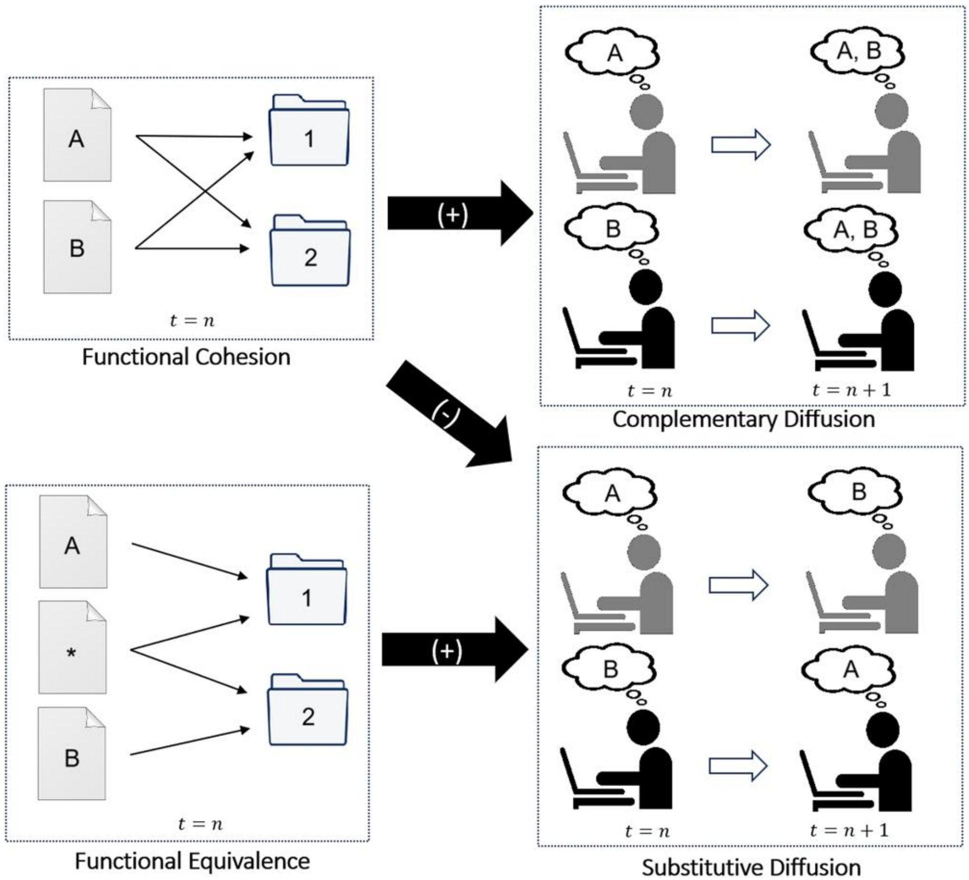

Fig. 1. An illustration of the two proposed types of functional relationships between languages (top-left and bottom-left) and the two proposed types of diffusion across users (top-right and bottom-right) and the (one-directional) hypotheses that exist between them (black arrows in center of figure). Functional cohesion (top-left) indicates that two languages (A and B) are featured in the same projects, whereas functional equivalence (bottom-left) indicates that two languages appear in the same projects with the same other languages (denoted by a "*"). Complementary diffusion (top-right) refers to the case where a user of A or B becomes a user of both A and B, whereas substitutive (or competitive) diffusion (bottom-right) refers to the case where a user of A becomes a user of B, or a user of B becomes a user of A.

across the minds of individual users. Our main set of independent variables are the overlaps between languages by project, and our main outcome variables are the overall usage and co-usage patterns of language by individual users. We use temporal lagging to ensure that the co-diffusion patterns across users are not determined by project co-usage.

More specifically, to test our hypotheses, we use Multiple Regression Quadratic Assignment Procedure (MR-QAP) to assess how functional cohesion and functional equivalence influence diffusion. The outcomes of interest are conceptually similar to what is estimated in an entry model (in the case of complementary diffusion) or an exit model (in the case of substitutive diffusion), but the key unit of observation are pairs of languages (technologies) in a connected network, as opposed to individuals or entities that are beginning or ceasing to use the technology themselves.

For all languages that hit a minimum usage threshold in year t, we calculate matrices (C and E) measuring functional relationships between each pair of languages at time t. We also construct matrices (U and V) estimating rates of diffusion between times t and $t+1$. A functional cohesion matrix (C) is calculated from the ratio of observed-to-expected cases of co-occurrence between each pair of languages. A functional equivalence matrix (E) is derived by calculating the cosine similarity of each pair of language vectors in the functional cohesion matrix, C. A complimentary diffusion matrix (U) is calculated by looking at the log-odds increase that using a given language A in year t had on newly adopting B in year $t+1$. A competitive/substitutive diffusion matrix (V) is calculated by looking at the log-odds increase that using a given language A in year $t+1$ had on ceasing usage of language B in year $t+1$. The final set of models uses MR-QAP to estimate the relationship between the functional matrices C and E have on the diffusion matrices U and V.

Below, we summarize our methodological strategy for calculating (i) functional relationships between languages, (ii) rates of diffusion, and (iii) estimating the latter as a function of the former.

Calculating functional cohesion and equivalence matrices

The GitHub dataset allows us to look at how frequently languages are being used together within projects, and how pairs of languages are used for similar purposes across projects (denoted by their use alongside similar other languages). "Functionally cohesive" languages are used together and frequently appear in the same projects, whereas "functionally equivalent" languages are used alongside the same other languages, suggesting they are interchangeable.

The entries in the matrix measuring functional cohesion, C, can be defined in terms of a measure of cohesion, c, for each pair of languages:

$$c_{A,B} = \sum_{i=1}^{n} \frac{I(A) * I(B)}{k_i * (k_i - 1) * \frac{1}{2}}$$

$$C_{A,B} = \log\left(\frac{c_{A,B}}{E[c_{A,B}]}\right).$$

Where A and B are row and column entries corresponding to specific programming languages, i is an index for each of the n projects from the year t, and k_i is the number of languages that are featured in project i. $I(A)$ and $I(B)$ are indicator variables for the presence of languages A and B within project i. The denominator, $k_i*(k_i-1)*\frac{1}{2}$, ensures that each project i contributes the same overall weight to the matrix. This is a way of projecting a set of hyperedges (51) that exist between one language and several projects into a weighted single-mode network of languages. The expected value of each cohesion value, $E[c_{a,b}]$, is calculated via simulated "re-shuffles" of the relationships between projects and languages, where languages occur in the same number of projects and each language's set of corresponding projects has the same distribution of "k" values (total number of languages). Borrowing from the methodological logic used to calculate chi-square values in the comparison of theoretical and observed distributions (52), we look at the ratio of the observed functional cohesion to what would be expected by chance. The logic also resembles the revealed comparative advantage (RCA) in the economic literature, which compares how much a certain nation exports a certain product compared to an expectation based on the nation's overall productivity and the overall global production of that product (53). Similar calculations have been used in more recent examinations of relatedness (23, 25, 54).

The cohesion matrix, C, is then used to create a functional equivalence matrix, E, that considers how languages A and B co-occur with each other language X:

$$E_{A,B} = \sum_{k} C_{A,X} * C_{B,X} * \frac{1}{\sqrt{\sum_X C_{A,X}^2}} * \frac{1}{\sqrt{\sum_X C_{B,X}^2}}.$$

Calculating diffusion rate matrices

The dependent variable matrices are calculated by looking at how many individuals use and do not use pairs of languages during two consecutive time windows. Complementary diffusion is determined by looking at how users of language A during time window t adopt language B compared to the likelihood of a user of neither language during time window t beginning to use language B during time window $t+1$. The cells of the matrix U are populated by measures of complementary diffusion from language A to B:

$$U_{A,B} = \frac{P(B_{t+1} \mid (A_t \wedge \bar{B}_t))}{P(B_{t+1} \mid (\bar{A}_t \wedge \bar{B}_t))} = \frac{\frac{|B_{t+1} \cap (A_t \cap \bar{B}_t)|}{|A_t \cap \bar{B}_t|}}{\frac{|B_{t+1} \cap (\bar{A}_t \cap \bar{B}_t)|}{|\bar{A}_t \cap \bar{B}_t|}}.$$

Where A_t and B_t refers to the set of users of language A and B respectively during time window t, and A_{t+1} B_{t+1} refers to the users of language A and B during time window $t+1$.

Substitutive (competitive) diffusion is determined by looking at the proportion of users of language A and B during time window t that stop using language B during time window $t+1$, compared to those that only used language B during time window t but did not use the language during time window $t+1$.

$$V_{A,B} = \frac{P(\bar{B}_{t+1} \mid (A_t \wedge B_t))}{P(\bar{B}_{t+1} \mid (\bar{A}_t \wedge B_t))} = \frac{\frac{|\bar{B}_{t+1} \cap (A_t \cap B_t)|}{|A_t \cap B_t|}}{\frac{|\bar{B}_{t+1} \cap (\bar{A}_t \cap B_t)|}{|\bar{A}_t \cap B_t|}}.$$

We limit our analysis to users who participated on the platform in both years for any given sequence of 2 years contained in our dataset, ignoring users who either enter the community in year $t+1$ or leave the community after year t. For each pair of adjacent time windows t and $t+1$, a randomly selected subset of 250,000 users who participated in a project in both years (or all users when the total number is less than 250,000) are analyzed. Each pair of languages that has a representation of higher than 0.5% in the first year of the sample are considered. (Separate models are eventually estimated for representation cut-off levels of 0.5, 1, and 2%.)

To limit the influence of outliers and normalize the variable distributions before implementing each of these dependent variables in the final set of models, the numerator and denominator of each side are given n-plus-one smoothing, and then the log of the overall ratio is taken. This makes the final formulas for the entries of matricies U and V:

$$U_{A,B} = \log\left(\frac{\frac{|B_{t+1} \cap (A_t \cap \bar{B}_t)| + 1}{|A_t \cap \bar{B}_t| + 1}}{\frac{|B_{t+1} \cap (\bar{A}_t \cap \bar{B}_t)| + 1}{|\bar{A}_t \cap \bar{B}_t| + 1}}\right);$$

$$V_{A,B} = \log\left(\frac{\frac{|\bar{B}_{t+1} \cap (A_t \cap B_t)| + 1}{|A_t \cap B_t| + 1}}{\frac{|\bar{B}_{t+1} \cap (\bar{A}_t \cap B_t)| + 1}{|\bar{A}_t \cap B_t| + 1}}\right).$$

The distribution of both functional predictors, compared to the two diffusion outcomes of interest, is shown in Figs. 2 and 3 for the year 2014 (with diffusion being measured from 2014 to 2015). Each point in the plots corresponds to one pair of languages, with functional equivalence plotted on the horizontal axis, functional cohesion plotted on the vertical axis, and the color corresponding to the observed level of diffusion between the two languages.

MR-QAP modeling of diffusion rates

While visualizations can shed some light on the nature of the data and relationships between variables, the estimation of network-level models of diffusion is necessary to contend with and account for the issues of dyadic independence that can confound the results of the relationship between one network and another.

A series of MR-QAP models are estimated for each pair of consecutive years (55, 56). This is a permutation-based procedure that allows the assessment of point estimates of the relationships between predictor and outcome networks. (For other more recent examples of how a series of how MR-QAP can be used to reveal insight across multiple network datasets, see Refs. (57) and (58).)

Here, this method illustrates how networks of normalized functional cohesion (*C*) and functional equivalence (*E*) between pairs of languages in year *t* shape networks of complementary diffusion (*U*) and substitutive diffusion (*V*) from year *t* to year $t+1$. Stated in terms of the matricies introduced in the prior sections, our two final models take the following form:

$$U_{t,t+1} = C_t + E_t$$
$$V_{t,t+1} = C_t + E_t.$$

From 2010 to 2018, models are estimated for complementary and substitutive diffusion that use networks of cohesion and equivalence. Separate models are estimated for each of the eight full 2-year sequences available in the data. The tests are permutation-based; from a hypothesis-testing perspective, we are primarily interested in how the coefficients generated for each network compare to the distribution of coefficients

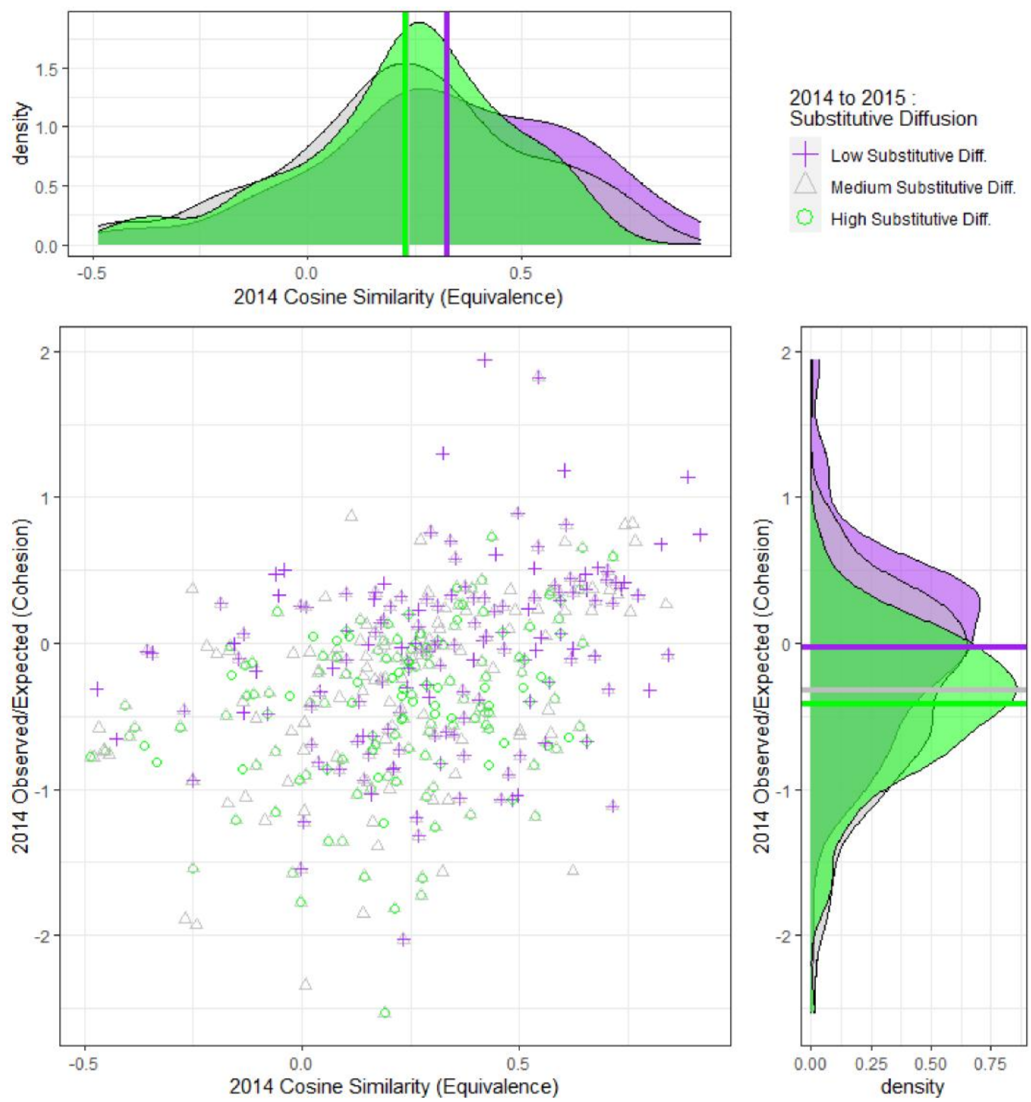

Fig. 2. This image shows the distribution of rates of functional equivalence, as measured by cosine similarity between languages, and functional cohesion, as measured by overlap in user bases, in relation to the overall observed levels of substitutive diffusion. Circular (green) points/density curves correspond to the 25% of language pairs with the highest levels of substitutive diffusion, cross (purple) points/density curves correspond to the 25% of language pairs with the lowest levels, and triangles (gray) correspond to the middle 50%. (Each (x,y) coordinate that corresponds to a pair of languages *A* and *B* also corresponds to the pair of languages *B* and *A*, so each point is slightly translucent to allow both points some visibility.)

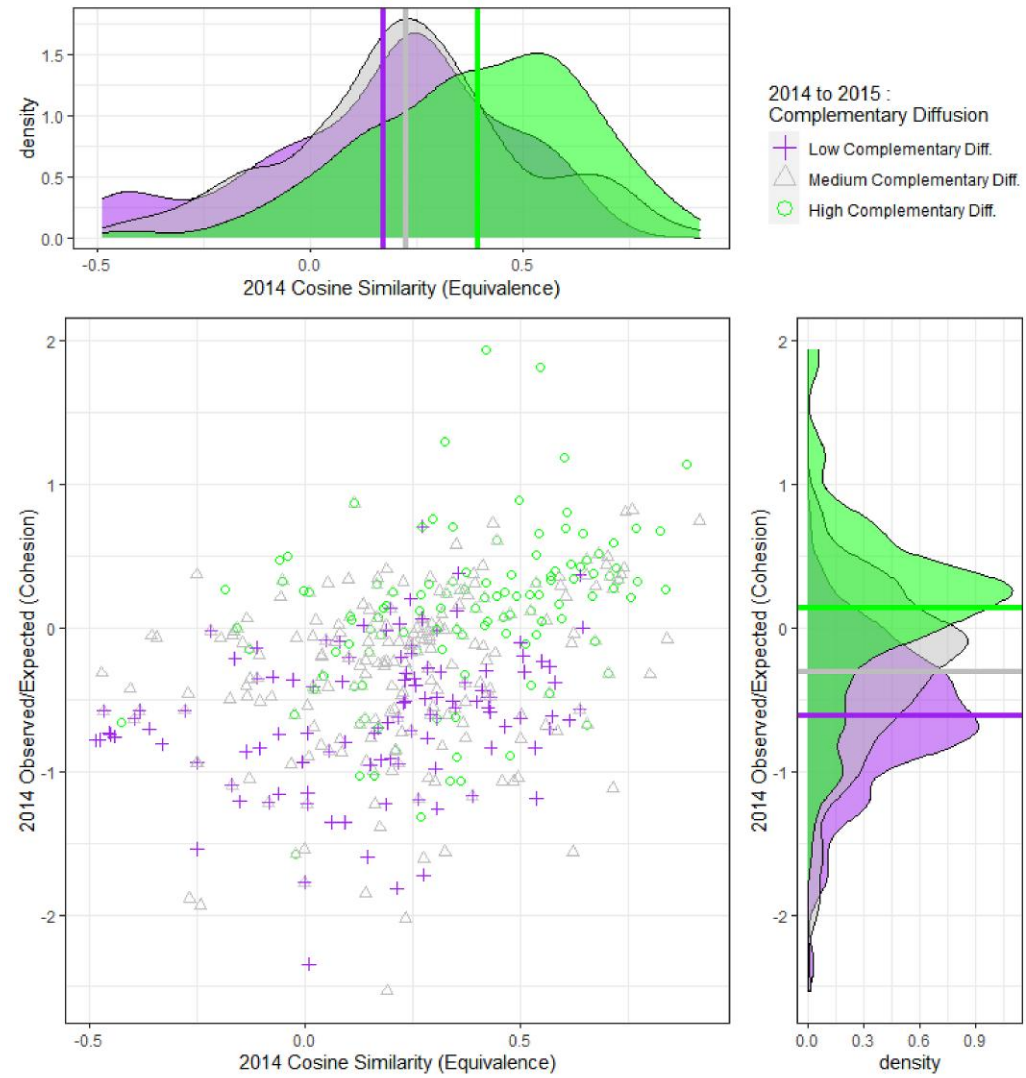

Fig. 3. This image shows the distribution of rates of functional equivalence, as measured by cosine similarity between languages, and functional cohesion, as measured by overlap in user bases, in relation to the overall observed levels of complementary diffusion. Circular (green) points/density curves correspond to the 25% of language pairs with the highest levels of complementary diffusion, cross (purple) points/density curves correspond to the 25% of language pairs with the lowest levels, and triangles (gray) correspond to the middle 50%. (Each (x,y) coordinate that corresponds to a pair of languages *A* and *B* also corresponds to the pair of languages *B* and *A*, so each point is slightly translucent to allow both points some visibility.)

generated by chance, as opposed to the overall explanatory power of the two variables.

There are three models for each year that correspond to different cut-off rates for language inclusion. Rates that are too high and exclusive limit analysis to a small number of languages, limiting potential statistical power, whereas rates that are too low will include low-popularity languages whose measures may add too much noise to the overall model. Models using rates of 0.5, 1, and 2% are all estimated to illustrate how this cut-off level affects the overall results and as a robustness check.

Network diagrams highlighting the observed relationship between each of the independent variables (functional cohesion and functional equivalence) and dependent variables (complementary diffusion and substitutive [competitive] diffusion) are presented in Figs. 4 and 5. Although these diagrams show the 2014–2015 time period with a cutoff 2%, our final models consider all 24 (eight 2-year periods * 3 cut-off levels) possible year and cut-off combinations contained in the dataset.

Results

We hypothesized that functional equivalence would positively predict substitutive diffusion, and that functional cohesion would positively predict complementary diffusion but negatively predict substitutive diffusion. We do not have a one-directional prediction for how equivalence may influence complementary diffusion, as mechanisms for both positive and negative effects both seem plausible. Table 2 shows a set of coefficients for six MR-QAP models for one year and cut-off level, the two complete models featuring both predictor networks for each outcome network, and four models using one predictor network and one outcome network.

The results of all 48 complete models featuring both predictor networks (8 years * 3 cut-off values * 2 types of diffusion) are shown in Fig. 6 and numerical values of the coefficients and *P*-values are provided in Tables 3 and 4. Our hypotheses on the role of functional cohesion are confirmed in both sets of models, with largely consistent results across both year and size of the cut-off. The co-occurrence of two languages in time *t* strongly and

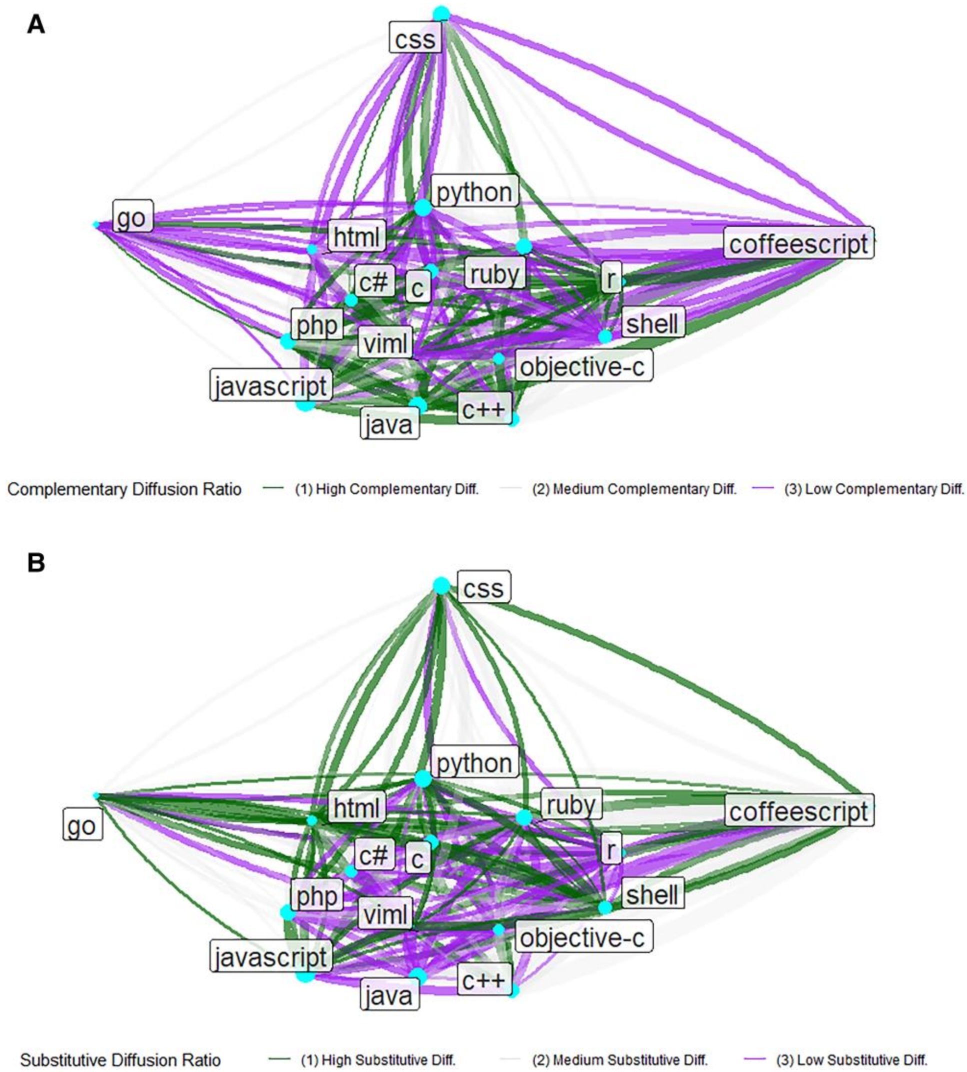

Fig. 4. Network of language co-occurrence and diffusion rates from 2014 to 2015, for languages that are used by over 2% of GitHub users. Edge thickness corresponds to the absolute value of the log observed-to-expected ratio, and edge color corresponds to whether the rate of diffusion is in the bottom 25%, middle 50%, or top 25% for the pair of languages. Complementary diffusion is shown on the top, and substitutive (competitive) diffusion is shown on the bottom.

positively predicts the tendency for users of one language to adopt the other language in time $t+1$. Furthermore, this negatively predicts the tendency for users of one language to abandon the other language at time $t+1$. This finding is statistically significant in 21 of 24 models at a level of $P < 0.05$.

The role of functional equivalence is more difficult to determine. Recall that there is not a one-directional prediction for how functional equivalence will be associated with complementary diffusion. There are reasonable mechanisms that could cause this relationship to be positive or negative. In models corresponding to earlier years in our data, equivalence tends to positively predict complementary diffusion at levels either at or near statistical significance, yet this effect fades over time. (This could be evidence of a mechanism that generates a negative relationship gradually becoming more powerful relative to a mechanism that generates a positive relationship, but this cannot be said with certainty.)

In models predicting substitutive (competitive) diffusion, functional equivalence tends to be a positive predictor, but only at a statistically significant level in 2% cut-off models for 2015 and 2016, while nearing positive statistical significance in the 2% models in 2010 and 2013. While the results in aggregate tend to suggest that this effect is more positive than expected by chance, especially in the 2% cut-off models that are less influenced by low-popularity languages, the predictor does not show the strength that functional cohesion does. Taken together, these results offer far milder support for the hypothesis that functional equivalence positively predicts substitutive diffusion in this empirical context.

Discussion

Building on prior examinations of "relatedness" and interacting contagions, the data and analysis demonstrate that the functional compatibility of diffusive phenomena can influence how they share and swap potential adopters. Prior work that draws on the concept of "product space" emphasizes how groups of actors in certain areas of this space are better poised to create new products or ideas. Here, we have focused on the fortunes of the creations, rather than the creators. This shift in focus can be especially informative when physical infrastructure (i.e. the presence of

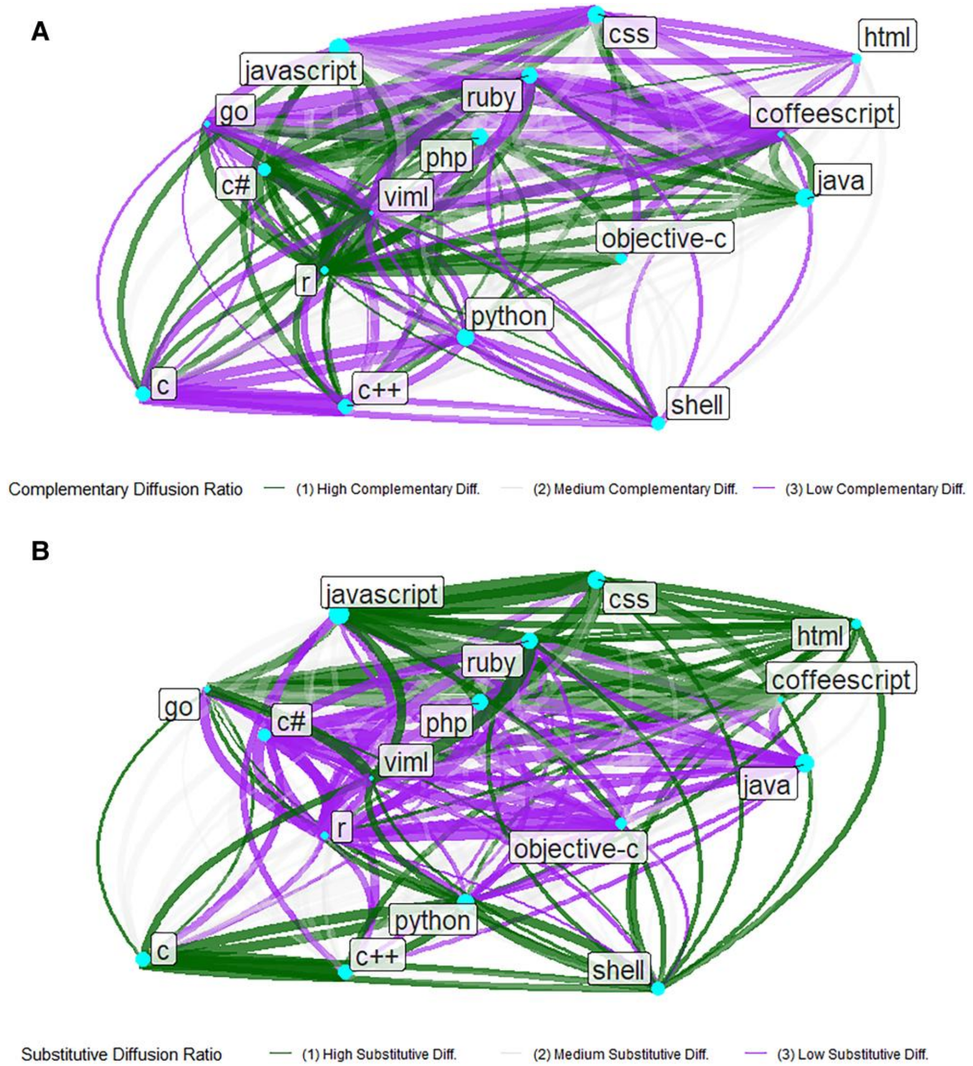

Fig. 5. Network of language equivalence and diffusion rates from 2014 to 2015, for languages that are used by over 2% of GitHub users. Edge thickness corresponds to the absolute value of the calculated equivalence between each pair of networks, and edge color corresponds to whether the rate of diffusion is in the bottom 25%, middle 50%, or top 25% for the pair of languages. Complementary diffusion is shown on the top, and substitutive (competitive) diffusion is shown on the bottom.

Table 2. Coefficients for MR-QAP models using networks constructed from the years 2014 and 2015 with a 2% cutoff for network inclusion (the same data shown in Figs. 4 and 5).

Models	**Complementary diffusion (*U*)**			**Substitutive diffusion (*V*)**		
Year: 2014–2015; Cutoff: 2%	**1**	**2**	**3**	**4**	**5**	**6**
Constant	0.4161	0.3118	0.4166	−0.1305	−0.1112	−0.1335
Network predictors						
Functional cohesion (C)	0.447		−0.0024	−0.0905		0.016
	(0.992)		(0.989)	(0.008)		(0.001)
Functional equivalence (*E*)		0.1848	0.4477		−0.0238	−0.0953
		(0.968)	(1.000)		(0.146)	(0.000)

For each outcome network (*U* and *V*), one model is shown with the cohesion matrix (*C*) as a predictor, one is shown with the equivalence matrix (*E*) as a predictor, and one is shown with both. *P*-values (in this context, the percentage of simulations that have a coefficient value that is less than the observed value) are shown in parentheses.

specific factories and utilities, the proximity of an appropriate workforce) is less of a limiting factor. Instead, the only limiting factor is the limited amounts of human attention and mind-share in a population (59). Coding languages are tools that can easily diffuse to new users, or disappear from their repertoire. Examinations of how sets of cultural or ideological phenomena become collectively correlated and anticorrelated in a population have illustrated the importance of socio-cognitive processes in the life cycle of ideas, beliefs, norms, or practices, regardless of their underlying content (60–62). Our analysis shows that the relationship between cognitive objects of diffusion does influence how they might spread with and against one another, even if

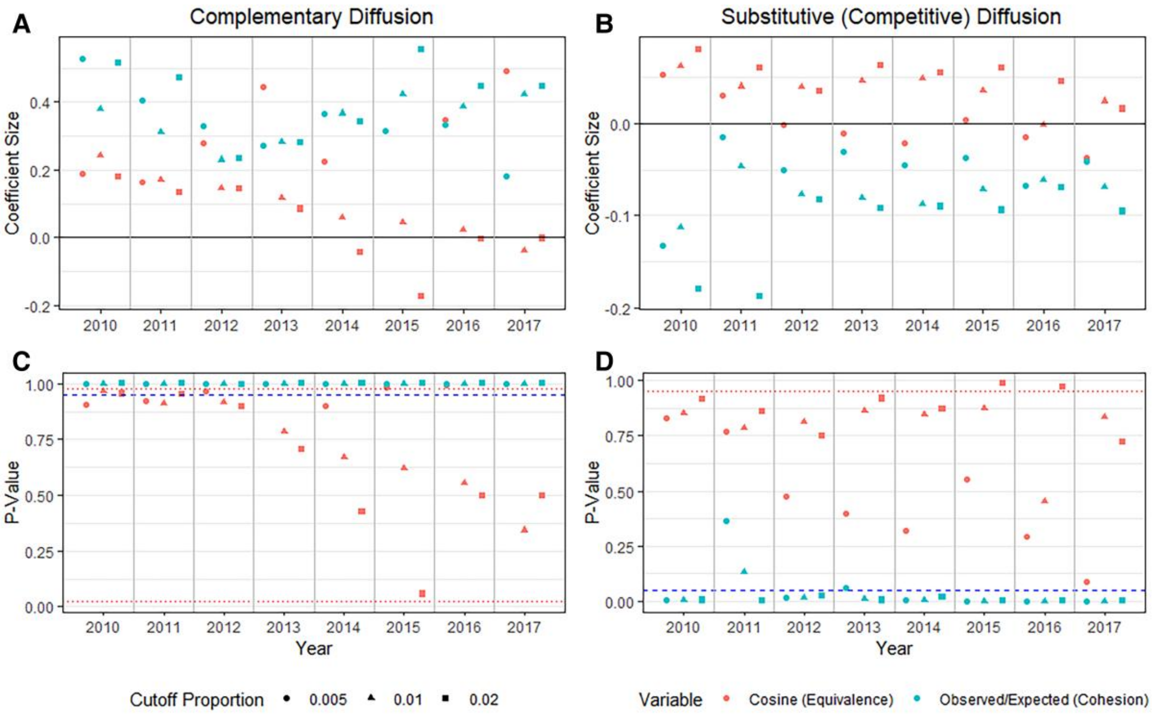

Fig. 6. Coefficients and P-values for each of the 24 models estimated for both complementary and substitutive (competitive) diffusion. Three models are estimated for cut-off points of 0.02, 0.01, and 0.005 for each year. Cohesiveness is a consistent positive predictor of complementary diffusion and a consistent negative predictor of substitutive diffusion. Functional equivalence, on the other hand, is a positive but somewhat inconsistent predictor of complementary and substitutive diffusion. In the case of complementary diffusion, the effect of functional equivalence shifts from positive to neutral over time, and in the case of substitutive diffusion, the effect of functional equivalence tends to be more positive with larger cut-off values.

Table 3. The results of MR-QAP models of complementary diffusion of languages between users over the course of two consecutive calendar years on the cohesion and equivalence of the languages in GitHub projects during the first year.

Cutoff	Year t	Year t+1	Constant	Cohes. Coef.	T-Stat	Sim.<Obs.	Equiv. Coef.	T-Stat.	Sim.<Obs.
0.005	2010	2011	0.254	0.53	7.714	1	0.188	2.179	0.907
0.005	2011	2012	0.262	0.406	6.743	1	0.165	2.141	0.92
0.005	2012	2013	0.338	0.329	6.783	1	0.28	3.522	0.964
0.005	2013	2014	0.286	0.272	7.275	1	0.445	5.799	0.997
0.005	2014	2015	0.348	0.367	8.698	1	0.226	2.431	0.9
0.005	2015	2016	0.398	0.314	7.121	1	0.317	4.034	0.984
0.005	2016	2017	0.548	0.334	7.477	1	0.348	4.235	0.994
0.005	2017	2018	0.432	0.181	5.303	0.998	0.491	6.666	1
0.01	2010	2011	0.215	0.382	5.482	1	0.242	2.657	0.966
0.01	2011	2012	0.227	0.311	4.817	1	0.17	2.079	0.909
0.01	2012	2013	0.226	0.23	4.441	1	0.145	2.154	0.917
0.01	2013	2014	0.272	0.283	4.863	0.998	0.116	1.481	0.784
0.01	2014	2015	0.248	0.367	5.803	0.998	0.06	0.751	0.67
0.01	2015	2016	0.326	0.422	7.399	1	0.043	0.484	0.623
0.01	2016	2017	0.394	0.386	8.09	1	0.022	0.251	0.557
0.01	2017	2018	0.424	0.422	8.912	1	−0.037	−0.516	0.344
0.02	2010	2011	0.085	0.516	6.068	1	0.18	1.935	0.955
0.02	2011	2012	0.079	0.472	6.049	1	0.132	1.66	0.95
0.02	2012	2013	0.196	0.235	4.089	0.994	0.143	2.003	0.895
0.02	2013	2014	0.246	0.28	4.317	1	0.084	0.976	0.706
0.02	2014	2015	0.221	0.341	4.742	0.999	−0.043	−0.465	0.426
0.02	2015	2016	0.212	0.556	8.076	1	−0.175	−1.976	0.057
0.02	2016	2017	0.425	0.448	7.8	1	−0.003	−0.035	0.498
0.02	2017	2018	0.417	0.448	8.494	1	−0.002	−0.038	0.496

While functional equivalence typically is a positive predictor of complementary diffusion, the association between functional cohesion and complementary diffusion shifts from positive to mixed over time.

relationships can emerge without compatibility. Accordingly, it shows how the principle of "relatedness" may be informative for continued studies of ideological and cultural diffusion in the social sciences. Just as compatible languages may help one another spread, the diffusion of philosophies, policies, and preferences may be shaped by their compatibility with other popular or "viral" ideas.

Accordingly, we hope the approach presented here can offer for guidance for how conceptions of product space can be extended

Table 4. The results of MR-QAP regressions predicting substitutive (competitive) diffusion of languages between users over the course of two consecutive calendar years on the cohesion and equivalence of the languages in GitHub projects during the first year.

Cutoff	Year *t*	Year *t*+1	Constant	Cohes. Coef.	T-Stat	Sim.<Obs.	Equiv. Coef.	T-Stat.	Sim.<Obs.
0.005	2010	2011	0.104	−0.133	−4.499	0.003	0.053	1.418	0.831
0.005	2011	2012	0.07	−0.014	−0.59	0.362	0.031	1.008	0.767
0.005	2012	2013	0.031	−0.05	−3.37	0.015	−0.001	−0.027	0.473
0.005	2013	2014	0.015	−0.03	−2.633	0.061	−0.01	−0.426	0.397
0.005	2014	2015	0.009	−0.045	−4.025	0.007	−0.022	−0.878	0.319
0.005	2015	2016	−0.033	−0.037	−4.167	0.001	0.004	0.252	0.551
0.005	2016	2017	−0.146	−0.068	−6.005	0	−0.015	−0.713	0.294
0.005	2017	2018	−0.111	−0.041	−4.402	0	−0.037	−1.832	0.087
0.01	2010	2011	0.088	−0.113	−3.774	0.004	0.062	1.591	0.854
0.01	2011	2012	0.055	−0.047	−1.857	0.132	0.041	1.277	0.788
0.01	2012	2013	−0.006	−0.076	−3.64	0.016	0.04	1.471	0.816
0.01	2013	2014	−0.018	−0.081	−4.354	0.009	0.047	1.865	0.863
0.01	2014	2015	−0.024	−0.088	−4.2	0.005	0.049	1.845	0.845
0.01	2015	2016	−0.036	−0.072	−5.548	0	0.035	1.732	0.874
0.01	2016	2017	−0.123	−0.061	−5.005	0	−0.002	−0.075	0.452
0.01	2017	2018	−0.124	−0.068	−4.834	0	0.025	1.139	0.838
0.02	2010	2011	0.085	−0.179	−3.945	0.005	0.081	1.62	0.915
0.02	2011	2012	0.101	−0.188	−3.992	0	0.06	1.253	0.862
0.02	2012	2013	−0.002	−0.083	−3.198	0.022	0.035	1.092	0.75
0.02	2013	2014	−0.015	−0.092	−4.217	0.005	0.063	2.193	0.919
0.02	2014	2015	−0.028	−0.09	−3.934	0.02	0.055	1.878	0.873
0.02	2015	2016	−0.025	−0.094	−5.765	0	0.061	2.887	0.987
0.02	2016	2017	−0.119	−0.07	−4.486	0	0.045	2.113	0.97
0.02	2017	2018	−0.133	−0.095	−5.87	0	0.016	0.838	0.724

The functional cohesion of languages negatively predicts substitutive diffusion, but functional equivalence tends to positively predict substitutive diffusion.

beyond the context of economic geography and incorporated into more studies of contagion, diffusion, and social influence. The methodological approach and findings presented can offer guidance for examining how other related sets of diffusive phenomena, be they products, practices, ideas, norms, beliefs, diseases, or fads, co-diffuse with one another. For example, Do political ideas that share philosophical compatibility spread with one another? Do different types of exercise supplant or complement one another in physical fitness routines? Do university courses on similar topics share or compete for enrollment? Do novel linguistic norms complement or challenge established modes of expression? We hope our approach can help inform future studies of network*ed* diffusions in the social sciences, network science, organizational behavior, marketing, communications, and science and technology studies.

More broadly, the choices, approaches, and adjustments taken here might provide guidance for future analyses of co-diffusion and bipartite network analysis. Much has been done by mathematicians, physicists, and computer scientists to find community and structure within bipartite networks (63–66) or to detect bipartivity in a network (67, 68), and social scientists have long been interested in using bipartite data to detect larger insights and patterns between people who share group things like group membership (69, 70), event attendance (40, 71, 72), or even purchase history (73). The approach presented here contributes to the continued development of such applied network research methods. For example, our approach offers a way for transforming what are effectively large sets of bipartite "hyperedges," which connect a node of one type (a project) to several nodes of another type (a language), to a bipartite projection which gives each hyperedge equal weight in the projection. With certain exceptions (74, 75), work on bipartite hypergraphs is extremely limited. Furthermore, the set of network data created provided a very rare opportunity to estimate MR-QAP models (for other examples, see Refs. (57, 58, 76)).

Conclusion

By conceptualizing diffusive phenomena in their own network schema, as opposed to independently considering their journeys through populations, we stand to gain a better understanding of how ideas and products spread with and against one another. By synthesizing the theory of relatedness and concept of product space with a network diffusion perspective, we hope to have shown how these two strands of research can enrich one another. While networks of potential adopters are of obvious importance when considering the spread of any given idea, norm, belief, or technology, network analysis can also account for how diffusive phenomena themselves are connected. Considering the contents and compatibilities of different ideas can shed light on more complex and multifaceted ecologies of technological change and innovation. This may shed new light not only on our understanding of not only technological diffusion but also polarization, disease, misinformation, and more complex social issues that can be better understood when incorporating the theory of relatedness and considering how larger sets of phenomena spread with and against each other. Ideas, products, innovations, and beliefs that are compatible with one another are likely to grow together.

Acknowledgments

The authors thank participants of the "Computational and Empirical Approaches to Micro-Macro Dynamics" session of the 2022 Meeting of the American Sociological Association for their comments and suggestions.

Funding

This work was supported by Natural Sciences and Engineering Research Council of Canada (NSERC #5660-501727) and Social Sciences and Humanities Research Council of Canada (SSHRC

#869-2016-0005), the Deutsche Forschungs- gemeinschaft (DFG; Germany #SCHR1282/3-1) and the National Science Foundation (NSF; United States #1723608) through the Trans-Atlantic Platform's Digging into Data Program.

Author Contributions

A.S.: conceptualization; data curation; formal analysis; investigation; visualization; methodology; writing original draft; writing-review and editing. J.M.: conceptualization; resources; data curation; software; visualization; methodology; writing-review and editing. N.Z.: resources; data curation; software; methodology; writing-review and editing. K.R.: supervision; funding acquisition; writing-review and editing. T.S.: supervision; funding acquisition; writing-review and editing.

Preprints

An earlier version of this manuscript was posted to the SocArxiv preprint server at https://doi.org/10.31235/osf.io/bfmy2.

Data Availability

The original data set from which the data in this article derived is from the GHTorrent project, which is currently documented at: https://github.com/ghtorrent/ghtorrent.org. The aggregated data used to construct this analysis is available on the Harvard Dataverse at https://doi.org/10.7910/DVN/CQVJTI. The code used to create results, tables, and visualizations from aggregated data is available on the corresponding author's GitHub repository at https://github.com/ads40788/InnovationCoDiffusion.

References

1 Rogers EM, Singhal A, Quinlan MM. 2014. *Diffusion of innovations*. Routledge.
2 Wejnert B. 2002. Integrating models of diffusion of innovations: a conceptual framework. *Annu Rev Sociol*. 28(1):297–326.
3 Centola D, Macy M. 2007. Complex contagions and the weakness of long ties. *Am J Sociol*. 113(3):702–734.
4 Guilbeault D, Becker J, Centola D. 2018. Complex contagions: a decade in review. In: Lehmann S, Ahn Y-Y, editors. *Complex spreading phenomena in social systems: influence and contagion in real-world social networks*. Springer. p. 3–25.
5 Cai W, Chen L, Ghanbarnejad F, Grassberger P. 2015. Avalanche outbreaks emerging in cooperative contagions. *Nat Phys*. 11(11): 936–940.
6 Hébert-Dufresne L, Althouse BM. 2015. Complex dynamics of synergistic coinfections on realistically clustered networks. *Proc Natl Acad Sci U S A*. 112(33):10551–10556.
7 Newman ME. 2005. Threshold effects for two pathogens spreading on a network. *Phys Rev Lett*. 95(10):108701.
8 Funk S, Gilad E, Watkins C, Jansen VA. 2009. The spread of awareness and its impact on epidemic outbreaks. *Proc Natl Acad Sci U S A*. 106(16):6872–6877.
9 Tai Z, Sun T. 2007. Media dependencies in a changing media environment: the case of the 2003 SARS epidemic in China. *New Media Soc*. 9(6):987–1009.
10 Bucklin LP, Sengupta S. 1993. The co-diffusion of complementary innovations: supermarket scanners and UPC symbols. *J Prod Innov Manage*. 10(2):148–160.
11 Arnold N. 2011. Technology diffusion of multiple simultaneous innovations. *Student Papers Complex Adaptive Systems Fall 2011*, p. 12.
12 Den Hartigh E, Van Asseldonk T. 2004. Business ecosystems: a research framework for investigating the relation between network structure, firm strategy, and the pattern of innovation diffusion. In: ECCON 2004 annual meeting: co-jumping on a Trampoline. The Netherlands.
13 Tomochi M, Murata H, Kono M. 2005. A consumer-based model of competitive diffusion: the multiplicative effects of global and local network externalities. *J Evol Econ*. 15(3):273–295.
14 Chang H-CH, Fu F. 2018. Co-diffusion of social contagions. *New J Phys*. 20(9):095001.
15 Vasconcelos VV, Levin SA, Pinheiro FL. 2019. Consensus and polarization in competing complex contagion processes. *J R Soc Interface*. 16(155):20190196.
16 Weng L, Flammini A, Vespignani A, Menczer F. 2012. Competition among memes in a world with limited attention. *Sci Rep*. 2(1):335.
17 Min B, San Miguel M. 2018a. Competing contagion processes: complex contagion triggered by simple contagion. *Sci Rep*. 8(1): 10422.
18 Min B, San Miguel M. 2018b. Competition and dual users in complex contagion processes. *Sci Rep*. 8(1):14580.
19 Myers SA, Leskovec J. 2012. Clash of the contagions: cooperation and competition in information diffusion. In: 2012 IEEE 12th International Conference on Data Mining. IEEE. p. 539–548.
20 Zhuang Y-B, Chen J, Li Z-H. 2017. Modeling the cooperative and competitive contagions in online social networks. *Phys A Stat Mech Appl*. 484:141–151.
21 Wang W, Liu Q-H, Liang J, Hu Y, Zhou T. 2019. Coevolution spreading in complex networks. *Phys Rep*. 820:1–51.
22 Hausmann R, Klinger B. 2006. Structural transformation and patterns of comparative advantage in the product space. Growth Lab Working Paper Series No. 5. p. 1–35.
23 Hidalgo CA, Klinger B, Barabási A-L, Hausmann R. 2007. The product space conditions the development of nations. *Science*. 317(5837):482–487.
24 Hausmann R, Hwang J, Rodrik D. 2007. What you export matters. *J Econ Growth*. 12:1–25.
25 Hidalgo CA, Hausmann R. 2009. The building blocks of economic complexity. *Proc Natl Acad Sci U S A*. 106(26):10570–10575.
26 Hidalgo CA. 2021. Economic complexity theory and applications. *Nat Rev Phys*. 3(2):92–113.
27 Hidalgo CA, *et al.* 2018. The principle of relatedness. In: Unifying Themes in Complex Systems IX: Proceedings of the Ninth International Conference on Complex Systems 9; Cambridge, MA. Cham: Springer. p. 451–457.
28 Kogler DF, Rigby DL, Tucker I. 2015. Mapping knowledge space and technological relatedness in US cities. In: *Global and regional dynamics in knowledge flows and innovation. Routledge*. p. 58–75.
29 Kogler DF, Essletzbichler J, Rigby DL. 2017. The evolution of specialization in the EU15 knowledge space. *J Econ Geogr*. 17(2): 345–373.
30 Castaldi C, Drivas K. 2023. Relatedness, cross-relatedness and regional innovation specializations: an analysis of technology, design, and market activities in Europe and the US. *Econ Geogr*. 99(3): 253–284.
31 Boschma R, Minondo A, Navarro M. 2013. The emergence of new industries at the regional level in Spain: a proximity approach based on product relatedness. *Econ Geogr*. 89(1):29–51.
32 Farinha T, Balland P-A, Morrison A, Boschma R. 2019. What drives the geography of jobs in the US? Unpacking relatedness. *Ind Innov*. 26(9):988–1022.

33 Hane-Weijman E, Eriksson RH, Rigby D. 2022. How do occupational relatedness and complexity condition employment dynamics in periods of growth and recession? *Reg Stud*. 56(7): 1176–1189.
34 Guevara MR, Hartmann D, Aristarán M, Mendoza M, Hidalgo CA. 2016. The research space: using career paths to predict the evolution of the research output of individuals, institutions, and nations. *Scientometrics*. 109:1695–1709.
35 Cheng M, *et al.* 2023. How new ideas diffuse in science. *Am Sociol Rev*. 88(3):522–561.
36 Kandler A, Steele J. 2008. Ecological models of language competition. *Biol Theory*. 3(2):164–173.
37 Valverde S, Solé RV. 2015. A cultural diffusion model for the rise and fall of programming languages. *Hum Biol*. 87(3):224–234.
38 Yun J, Shang S-C, Wei X-D, Liu S, Li Z-J. 2016. The possibility of coexistence and co-development in language competition: ecology–society computational model and simulation. *Springerplus*. 5(1):1–16.
39 Juhász S, Wachs J, Kaminski J, Hidalgo CA. 2024. The software complexity of nations, arXiv, arXiv:2407.13880. https://doi.org/10.48550/arXiv.2407.13880, preprint: not peer reviewed.
40 Breiger RL. 1974. The duality of persons and groups. *Soc Forces*. 53(2):181–190.
41 Burt RS. 1978. Cohesion versus structural equivalence as a basis for network subgroups. *Sociol Methods Res*. 7(2):189–212.
42 Lorrain F, White HC. 1971. Structural equivalence of individuals in social networks. *J Math Sociol*. 1(1):49–80.
43 Porter ME, Strategy C. 1980. *Competitive strategy: techniques for analyzing industries and competitors*. New York: Free Press.
44 Chen T, *et al.* 2020. Try this instead: personalized and interpretable substitute recommendation. In: Proceedings of the 43rd International ACM SIGIR Conference on Research and Development in Information Retrieval. Association for Computing Machinery. p. 891–900.
45 Wang Z, Jiang Z, Ren Z, Tang J, Yin D. 2018. A path-constrained framework for discriminating substitutable and complementary products in e-commerce. In: Proceedings of the Eleventh ACM International Conference on Web Search and Data Mining. Association for Computing Machinery. p. 619–627.
46 Hoey J, *et al.* 2021. Theoretical and empirical modeling of identity and sentiments in collaborative groups. SocarXiv. https://doi.org/10.31235/osf.io/4hukx, preprint: not peer reviewed.
47 Zöller N, Morgan JH, Schröder T. 2020. A topology of groups: what GitHub can tell us about online collaboration. *Technol Forecast Soc Change*. 161:120291.
48 Terrell J, *et al.* 2017. Gender differences and bias in open source: pull request acceptance of women versus men. *PeerJ Comput Sci*. 3:e111.
49 Vedres B, Vasarhelyi O. 2019. Gendered behavior as a disadvantage in open source software development. *EPJ Data Sci*. 8(1):25.
50 Vasilescu B, *et al.* 2015. Gender and tenure diversity in GitHub teams. In: Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing systems. Association for Computing Machinery. p. 3789–3798.
51 Berge C. 1984: *Hypergraphs: combinatorics of finite sets*. Vol. 45. Elsevier.
52 Pearson K. 1900. On the criterion that a given system of deviations from the probable in the case of a correlated system of variables is such that it can be reasonably supposed to have arisen from random sampling. *London Edinburgh Philos Mag J Sci*. 50(302):157–175.
53 Balassa B. 1965. Trade liberalisation and “revealed” comparative advantage. *Manchester Sch Econ Soc Stud*. 33(2):99–123.
54 Neffke F, Henning M. 2013. Skill relatedness and firm diversification. *Strateg Manag J*. 34(3):297–316.
55 Dekker D, Krackhardt D, Snijders TA. 2007. Sensitivity of MRQAP tests to collinearity and autocorrelation conditions. *Psychometrika*. 72(4):563–581.
56 Krackhardt D. 1988. Predicting with networks: nonparametric multiple regression analysis of dyadic data. *Soc Netw*. 10(4): 359–381.
57 Schon J, Johnson JC. 2021. How inter-state amity and animosity complement migration networks to drive refugee flows: a multilayer network analysis, 1991–2016. *PLoS One*. 16(1):e0245712.
58 Baird T. 2017. Who speaks for the European border security industry? A network analysis. *Eur Secur*. 26(1):37–58.
59 Simon HA, Deutsch KW, Shubik M. 1971. Designing organizations for an information-rich world. *Comput Commun Public Int*. 72:37.
60 DellaPosta D. 2020. Pluralistic collapse: the ‘oil spill’ model of mass opinion polarization. *Am Sociol Rev*. 85(3):507–536.
61 DellaPosta D, Shi Y, Macy M. 2015. Why do liberals drink lattes? *Am J Sociol*. 120(5):1473–1511.
62 Goldberg A, Stein SK. 2018. Beyond social contagion: associative diffusion and the emergence of cultural variation. *Am Sociol Rev*. 83(5):897–932.
63 Barber MJ. 2007. Modularity and community detection in bipartite networks. *Phys Rev E Stat Nonlin Soft Matter Phys*. 76(6):066102.
64 Dormann CF, Strauss R. 2014. A method for detecting modules in quantitative bipartite networks. *Methods Ecol Evol*. 5(1):90–98.
65 Guimera R, Sales-Pardo M, Amaral LAN. 2007. Module identification in bipartite and directed networks. *Phys Rev E Stat Nonlin Soft Matter Phys*. 76(3):036102.
66 Larremore DB, Clauset A, Jacobs AZ. 2014. Efficiently inferring community structure in bipartite networks. *Phys Rev E*. 90(1): 012805.
67 Chen L, Yu Q, Chen B. 2014. Anti-modularity and anti-community detecting in complex networks. *Inf Sci*. 275:293–313.
68 Holme P, Liljeros F, Edling CR, Kim BJ. 2003. Network bipartivity. *Phys Rev E*. 68(5):056107.
69 Benton RA. 2016. Corporate governance and nested authority: cohesive network structure, actor-driven mechanisms, and the balance of power in American corporations. *Am J Sociol*. 122(3): 661–713.
70 Neal Z. 2014. The backbone of bipartite projections: inferring relationships from co-authorship, co-sponsorship, co-attendance and other co-behaviors. *Soc Netw*. 39:84–97.
71 Davis A, Gardner BB, Gardner MR. 2009. *Deep south: a social anthropological study of caste and class*. University of South Carolina Press, 1941.
72 Everett MG, Borgatti SP. 2013. The dual-projection approach for two-mode networks. *Soc Netw*. 35(2):204–210.
73 Shi F, Shi Y, Dokshin FA, Evans JA, Macy MW. 2017. Millions of online book co-purchases reveal partisan differences in the consumption of science. *Nat Hum Behav*. 1(4):0079.
74 Chen H, Frieze A. 1996. Coloring bipartite hypergraphs. In: *International Conference on Integer Programming and Combinatorial Optimization*. Springer. p. 345–358.
75 Zverovich I, Zverovich I. 2006. Bipartite bihypergraphs: a survey and new results. *Discrete Math*. 306(8-9):801–811.
76 Park JH, Kwahk K-Y, Han H, Kim Y. 2013. A study on the determinants of patent citation relationships among companies: MR-QAP analysis. *J Intell Inf Syst*. 19(4):21–37.