

How Much Should You Pay to Speak Esperanto? Quantifying Cultural Impacts on Trade through Wikipedia Network Statistics

M.J. Histen^{*}, A. Yukhov[†]

Abstract: Culture shapes how populations interact, and groups with more similar cultures face less frictions to cooperate. This, in turn, affects trade patterns and institutional diffusion, implying a society's culture plays a crucial role in the causes of economic growth. However, operationalizing cultural traits are challenging. We develop a new systematic measure through network analysis of Wikipedia. Connectivity statistics over the encyclopedia's hyperlink-directed network capture unique features of cultural relatedness. Further, as we index pages, we can coarsen the network into specific topics (e.g., cuisine, language, fashion). We aim these cultural proxies at OECD country trading patterns. The results show how culture correlates broadly over trade pattern data, and sheds insight into the impact of variety of cultural dimensions on the cooperation between populations.

JEL Classifications: F1, O3, Z1

Keywords: institutions, culture, network analysis

^{*} matthew.histen@csun.edu; California State University, Northridge, Northridge, CA

[†] alex.yukhov@csun.edu; California State University, Northridge, Northridge, CA

Culture shapes how cooperation takes place along a variety of activities. Populations with similar cultures face less frictions to interacting and can more easily adopt new developments from ones similar to themselves. Thus, a society's culture plays a crucial role in the diffusion of productivity enhancing innovations fundamental to the causes of trade and long-run economic growth.

While many studies have attempted to empirically quantify relationships between trade and culture, cultural traits are challenging to characterize and measure. Most of these approaches use survey answers or experiments focusing on general attitudes related to trust, individualism, and perceptions of work and poverty (Alesina and Giuliano, 2015). But by pinning culture to explicitly stated preferences, these studies rely on idiosyncratic sources with results that are hard to generalize.

We contribute to this literature by developing a new systematic measure of culture through network analysis over Wikipedia. We derive cultural measures through statistics about article connectivity, an untapped data source in cultural research. Articles cross reference each other to create a directed network between topics. When aggregating network statistics across thousands of hyperlinks, we obtain a systematic measure of cultural relatedness between countries. Furthermore, because each page in Wikipedia is indexed by categories, we can disentangle similarity measures along different dimensions including religion, language, cuisine, and more. These measures are consistent with previous results across different source data including World Value Surveys, linguistic bifurcation graphs, religious genealogies, and others. Thus, Wikipedia data can be used to generate, separate, and compare cultural similarity across a range of dimensions, and statistical variations across them can remark on the relative reach of history. These data robustly correlate with a variety of OECD country trade patterns, supporting an argument of culture as crucial institution in guiding country interactions.

Related literature

Institutions are the “rules of the game” that standardize the patterns of interaction, encompassing formal rules and social conventions, written laws, informal norms, and shared beliefs (North, 1990, p. 3). Standardization, by definition, implies interactions conform to an expectation. Institutions define the form of such association (Hodgson, 2006). These rules provide in advance knowledge about how an interaction will proceed, providing a means for planning, predicting, and cooperating (Hayek, 2011, p. 268). Many researchers have asked which institutions survive given their context and behavior. The process of establishing the right institutions is endogenous (Chang, 2010; Rubin, 2017, p. 251). Success rests on the fact that an individual can benefit from more knowledge than she is aware of through products of cumulative growth in cultural adaptations (Hayek, 2011, p. 90).

The notion of culture to economists remains rather vague (Nunn, 2020a), and definitions of “institutions” often overlap with definitions of “culture” (Spranz, Lenger, and Goldschmidt, 2012; Alesina and Giuliano 2015). Still, empirical papers attempt to distinguish formal institutions (formal legal or regulatory systems) from informal ones (norms, beliefs, and informal rules), but these tend to be performed in isolation of each other (Alesina and Giuliano, 2015). Moreover, merely bucketing what's left after subtracting a legal apparatus as “informal institutions” or “culture” is both theoretically and empirically insufficient.

To avoid relegating either to secondary importance, we take a holistic view of culture as encompassing a society's formal and informal institutions affecting cooperation. That is, institutions are the rules of the game *and* how we play it. Cultural similarity aims at how easily populations can share knowledge, communicate, and interact. It is often unclear how to distinguish between formal and informal institutions as conventions and norms tend to become codified overtime, which feedback to affect the evolution of these conventions and norms (Nunn, 2020a). Moreover, informal institutions are often prerequisites to formal ones (e.g., democratic values are necessary for legalized democratic institutions to operate (Besley and Persson, 2019)). Ultimately, similar cultural institutions (including formal ones) influence the

diffusion of complex technological innovations (Spolaore and Wacziarg, 2016) and the facilitation of trade flows (Guiso, Sapienza, and Zingales, 2006).

However, culture is hard to pin down. Methodologically, a focus on culture struggles to bring to bear tools of quantitative methods to parse the characteristics of specific cultural traits. The quest for a summary measure of cultural differences is challenging since many cultural practices – food, dress, accent, etc. – are not collected in any systematic manner (Abramitzky, 2015). Most empirical papers represent culture as the broad attitudes and values related to generalized trust, individualism, and perceptions of work and poverty (Alesina and Giuliano, 2015). Typically, these values are measured through survey questions aggregated at the country level (e.g., the World Values Survey) and then correlated with economic outcomes. These studies calculate a culture trait through questionnaires and correlate it to an aggregate measure (e.g., more individualistic cultures are less wedded to tradition and less likely to show nepotism, which affects the organization of companies and politics and has consequences for long run growth (Gorodnichenko and Roland, 2011; Acemoglu and Robinson, 2012)).

Another line of research involves economic experiments. For example, participants from different cultures are briefed on a coordination game (e.g., the dictator game (Duhaime, 2015) or the ultimatum game (Henrich, 2000)) or a task with priming (e.g., unscramble sentences with words influenced by a studied topic such as religion (Norenzayan and Shariff, 2008)). Their results are again interpreted as an index of attitudes and values (e.g., individualism) which can be aggregated and regressed against different outcome variables. Studies show relationships across a range of phenomena ranging from cooperation with strangers, voluntary blood donations, or parking tickets for diplomats (Everett, Haque, and Rand, 2016; Henrich 2020, p. 212).

Clearly, these approaches detect intriguing relationships. But imposing a top-down structure of culture framed by predefined traits or attitudes constrains broader connotations of culture. These studies also tend to collapse culture into a single proxy like trust, individualism, or whether two countries have been to war with each other (e.g., Guiso, Sapienza, and Zingales, 2009). This often lacks generalizability since these proxies are unique and from noncomparable data sources, yet measures are aggregated and conclusions drawn across studies (Alesina and Giuliano, 2015). Furthermore, they miss crucial elements that are harder to represent as a regression variable, or aspects that haven't been thought of altogether. Elements like music, cuisine, fashion, etc. matter significantly in how populations interact but aren't easily quantified through surveys or experiments.

Quantitative methods over big data sources are beginning to offer more generalizability. Parsing large corpuses of text to investigate cultural ideas through dictionaries or topic models have been used to generate cultural proxies in regression models (Blaydes, Grimmer, and McQueen, 2018; Grajzl and Murrell, 2019). For example, Michalopoulos and Xue (2021) analyze cultural attitudes by analyzing topics from country specific folklore. We contribute to the study of culture by extending these big data analyses to networks. We argue that the network structure of Wikipedia provides a novel dimensional space to measure cultural relatedness. A country's Wikipedia page contains thousands of hyperlinks to other pages within the encyclopedia. These links reveal aspects of their cultural attention, and clusters around salient items are quantifiable.¹ When aggregated across thousands of links for each country, we

¹ For example, consider cultural activities through sports. Even if India's Wikipedia page does not contain a direct link to "baseball," there's a short path to that topic through other pages in the network (e.g., "cricket"). Properties about this distance provide a measurable closeness around different cultural items. In fact, to celebrate sixty years of diplomatic relations, in 2013 joint efforts by the Japanese and Indian governments created an anime series called "Suraj: The Rising Star." A reboot of a Japanese anime about baseball, its re-release in India centered around cricket and required few other changes. The leaders of the project argued that it synthesized Indo-Japanese values and strengthened economic, diplomatic, and cultural traditions across the two countries.

obtain a broad, systematic measure of cultural similarity between countries. Similarity is decentralized and emergent compared to textual analyses which involve more researcher judgement over which corpuses are included. Moreover, Wikipedia categorizes each page into topics. Separating by category tags, then, allows us to disentangle relationships along the lines of religion, language, cuisine, and more.

Overall, the body of work studying trade patterns increasingly emphasizes the role of ethnic, linguistic, religious, and other cultural effects (Michalopoulos and Papaioannou, 2016; Nunn 2020b; Michalopoulos and Xue, 2021). We hope to open the door to additional channels of cultural impact and unearth even deeper roots. Using Wikipedia, we generate results consistent with previous literature from a systematic method over a single data source and extend the analysis onto categories not yet considered. Thus, Wikipedia as data offers several advantages compared to previous work. Its decentralized, emergent structure allows for an extemporaneous capture of broad cultural traits versus predefined buckets. Its massive size stretches across many topics to allow for a broad and systematic characterization of culture, and its organizational characteristics allow for slicing the data along many dimensions. Wikipedia as data provides a novel measure of culture and connecting it to genealogy reinforces old answers and contributes new insights regarding critical phenomena in comparative economic development.

Data

Open-source collaboration involves a unique production model where users trade effort outside of traditional organizational forms like firms or markets (Langlois and Garzarelli, 2008). Provided the architecture for interaction is well-established, users select into activities suiting their specialization. Wikipedia is an open-source encyclopedia project based on a model of freely editable content. By taking advantage of dispersed and idiosyncratic knowledge, 125,000 active monthly contributors and 40 million total accounts (Wikistats, 2023) exchange their specialized knowledge currently. This has led to Wikipedia becoming far more comprehensive than any other encyclopedia. At the time of analysis, it contained more than fifty-seven million articles across three hundred languages, with over six million articles written in English alone (Wikistats, 2023).

Because it's written collaboratively by anonymous volunteers, anyone can edit its media, leaving articles prone to misinformation, errors, or vandalism. These concerns vary substantially across language versions and articles.² Reliability is by far most consistent in the English version (Steinsson, 2023), which populates the data analyzed here. In the English Wikipedia, several mechanisms are in place to maximize contributions while preserving quality. Articles are standardized through an established manual of style, both for ordering of sections and formatting. The site tries to operate under strict policies of viewpoint neutrality and information verifiability. Aiming to be only a tertiary source, Wikipedia is explicit about not performing original research and instead acting entirely on previously published information, requiring material to be verified through qualifying sources with direct inline citation. Sources must also be publicly available in some form. When claims are exceptional, multiple high-quality sources are required.

Wikipedia operates by open and transparent consensus for editorial disputes, too, with well-defined processes for resolution. Editorial control falls on a network of self-organizing editors subject to peer review. These administrative roles are only achieved by nomination, which helps enforce communal standards of conduct backed by experience and familiarity with its content policies. Thousands of active editors will be using, monitoring, or editing the articles and assisted by automated programs to help watch for problematic edits (Wikistats, 2023). Popular or controversial articles are reviewed more closely to make it harder for vandalism. An arbitration committee sits on top of all editorial disputes, composed of

² Note that language versions of the encyclopedia are not translations of each other. Rather, contributors of each language generate their own articles. This means articles can vary substantially between versions, if they exist across them at all.

elected members in regularly rotated tranches. This committee has the authority to impose binding solutions that the community struggles to resolve on its own.

Because Wikipedia has so many users, articles are scrutinized by a massive number of readers and editors, with every edit logged and every debate about edits logged. The fact that anyone can edit articles increases the chances that errors will be corrected rather than propagated. Reliability of content has been analyzed many times. Though results vary, project outcomes are generally positive (Mesgari et al., 2015; Petiška and Moldan, 2019; Steinsson, 2023). Since 2005, Wikipedia has shown an error rate similar to Britannica (Giles, 2005) with most vandalism resolved within five minutes (Viégas, Wattenberg, and Dave, 2005). Though most contributors are deeply committed to curating accurate information according to their dispersed knowledge, criticisms are justified over content errors and editorial processes. There have been many examples of false or misleading information persisting for months, a systematic political bias (Greenstein and Feng, 2012), and demographic concerns related to the composition of editors who skew heavily male from Western countries (Meyer, 2013). Given that there is no systematic process to generate “obviously important” articles, they are populated by the relative attention of its users’ linguistic and cultural interests.

Can the English Wikipedia be trusted, then? The short answer is, on average, yes, and accuracy improves substantially for popular articles. Popular articles, by the nature of network analysis, populate most of the data in this study. Even if we discount each article’s efficacy, the unit of analysis is not the content itself but the network of hyperlinks cross-referencing articles. In fact, this structure is even more emergent than the generation of articles and harder to intentionally manipulate. Ultimately, Wikipedia is a widely contributed platform aggregating knowledge from many different viewpoints, offering a unique insight into different cultures’ attention and priorities.

Data Collection

Wikipedia as data involves representing the robust network of hyperlinks between pages as a mathematical object. All articles are linked or cross-referenced throughout the encyclopedia to guide users to related pages with additional information. This network, then, can be transformed into a directed graph. As part of its manual of style, contributors are encouraged to link to other articles with relevant connections regarding background information, events, proper nouns, technical information, and more (Wikistats, 2023). Downward pressure exists to avoid overlinking to pages considered commonly understood. These pathways form a network structure that connects topics organically. To derive this network, we began with 207 twentieth century countries from OECD.org and scraped all internal hyperlinks that appear in paragraph html. After removing redundant, administrative, and miscellaneous pages, each country had on average 523 content links to other material articles. Italy had the most with 1,700 while Democratic Republic of Congo had the fewest with 100 links. Since the analysis is over the English Wikipedia, Western countries receive more attention by contributors. Further work can clarify these differences across language versions of the encyclopedia. Collecting these country page links summed to about 70,000 unique pages, which were also scraped for all their internal hyperlinks appearing in paragraph html. This generated an adjacency matrix of about 1.5 million unique articles.

Simultaneously, all category tags from each page were scraped. Wikipedia is so vast that several features have been introduced to facilitate navigation, including categorizing every page by topics. Though categorization efforts are continuous, the bottom of every page contains relevant tags that classify pages into larger groupings. Each category can be browsed for related pages clustered under the same tag to create an interconnected hierarchy. Like the content itself, categorization develops collaboratively by end users applying public tags. Wikipedia’s general taxonomy contains about forty irreducible topics.³ After

³ This list contains {“Academic disciplines”, “Business”, “Concepts”, “Culture”, “Economy”, “Education”, “Energy”, “Engineering”, “Entertainment”, “Entities”, “Ethics”, “Events”, “Food and drink”, “Geography”, “Government”, “Health”,

collecting tags for all articles in the matrix, we can coarsen the network into specific categories (e.g., drop all pages not related to a topic). Table 1 gives more detailed information on the coarsened categories, which Wikipedia main topics fall under their umbrellas, and additional information related to the types of pages.

Table 1: Coarsened network criteria by Wikipedia Main Topic Classifications.

Regression Variable	Wikipedia Main Topic Classifications	Sample Pages	Network Size
business	{ "Business", "Economy", "Industry" }	{ "World Trade Organization", "Bloomberg L.P.", "Lomé Convention", "Kmart", "Jollibee", "Outsourcing", "Tender board", "Bretton Woods Conference", "Regional Comprehensive Economic Partnership", "Market capitalisation" }	68,565
culture	{ "Culture", "Entertainment", "Food and Drink", "Music", "Sports" }	{ "Tinikling", "Steinstossen", "Iranian art", "Beef Stroganoff", "Asphyx", "Hua-Yi distinction", "Batwing sleeves", "Pakol", "Walkabout long Chinatown", "The Adventures of Pinocchio" }	297,572
education	{ "Education" }	{ "College-preparatory school", "Universitas 21", "Bocconi University", "Art schools", "Licentiate (degree)", "Science, technology, engineering, and mathematics", "Direct Subsidy Scheme", "The three Rs", "Catholic school" }	29,195
events	{ "Events" }	{ "Hurricane Hanna", "1896 Summer Olympics", "1992 Cannes Film Festival", "March 23 Movement", "1815 eruption of Mount Tambora", "Expo 2020", "Southern Movement", "2012 ICC World Twenty20", "Kuwaiti oil fires", "Epiphany (holiday)" }	196,272
government	{ "Government", "Politics" }	{ "National personification", "Commonwealth republic", "Royalism", "Dwifungsi", "Ministry of Interior", "Carousel voting", "Arengo", "Supermajority", "Wealth tax", "Dutch Sandwich" }	382,030
language	{ "Language" }	{ "Ghanaian Pidgin English", "Ie (digraph)", "Deutsches Wörterbuch", "Adyghe language", "Jopara", "Ishkashimi language", "Ems Ukaz", "Qatari Sign Language", "Tajik alphabet", "Cursive script (East Asia)" }	92,919
military	{ "Military" }	{ "Brigade (military)", "Battle of Salamis", "Balkan Wars", "Conscription", "Nuclear proliferation", "Fieldcraft", "Lajes Air Base", "V-2 rocket", "Biological Weapons Convention", "Conscientious objection" }	88,618
people	{ "People" }	{ "Alaric I", "Marco Polo", "Aung Myint", "Al Ali (tribe)", "Kven people", "Riri Riza", "Athol Fugard", "Hun Sen", "Salian dynasty", "Roma in Greece" }	221,289
religion	{ "Religion" }	{ "Saltigue", "Stations of the Cross", "Ivion monastery", "Millenarianism", "Shwedagon Pagoda", "Neoplatonists", "Hindu calendar", "Temple of the Tooth", "Ise Grand Shrine", "Islamic cosmology" }	86,458
technology	{ "Engineering", "Mathematics", "Technology" }	{ "Stabilizer (aircraft)", "Satellite dishes", "Transistor", "Welding", "Nuclear power plants", "Materials science", "Semiconductor fabrication plant", "Vernacular architecture", "Stem cell", "Wireless internet" }	62,056

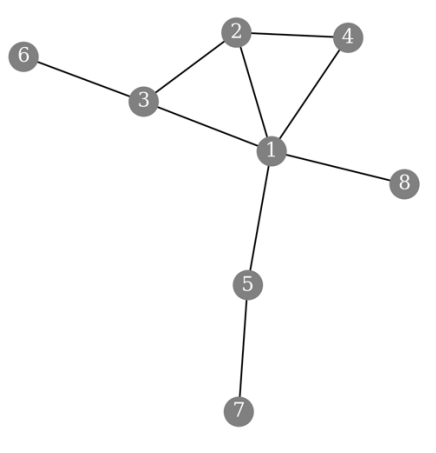
Dependent variable

Using tools from network analysis, we move from qualitative concepts of cultural similarity to quantitative measures through clusters in the network. A network simplifies a system to an abstract structure preserving only its connections (Newman, 2010, p. 105). These connections describe the paths along which something flows. Centrality statistics are real-valued measures of importance along this interpretation, where values provide rankings of critical nodes (Newman, 2010, p. 160). Applications of centrality include super-spreaders of disease or key infrastructure in transportation networks. One of the most common is degree centrality. The degree of a node refers to the number of connections it has to other nodes, which can be interpreted as a measure of likelihood for a node to catch flows through the network. Degree centrality, then, considers the number of links incident upon a node to rank its relative importance by weighting the number of walk lengths across the network (Newman, 2010, p. 160). For example, consider the network shown in Figure 1. The node with the highest degree centrality would be “1” which is a connection point for most information flowing through the network, while node “7” would have lowest degree centrality. For Wikipedia, the network forms from cross referenced hyperlinks between articles and flow captures how information propagates through nodes. Table 2 below shows articles with the highest degree centrality of the different subnetworks.⁴

“History”, “Human behavior”, “Humanities”, “Industry (economics)”, “Information”, “Knowledge”, “Language”, “Law”, “Life”, “Mass media”, “Mathematics”, “Military”, “Music”, “Nature”, “People”, “Philosophy”, “Politics”, “Religion”, “Science”, “Society”, “Sports”, “Technology”, “Universe”}.

⁴ Because the network was grown from country article links, it filled out along high degree nodes from that starting point.

Figure 1: A force-directed graph drawing of a randomly generated network.



Our dependent variable comes from a network statistic that effectively inverts centrality measures: connectivity. Connectivity measures the minimum number of paths that need to be removed before two nodes are no longer connected (Newman, 2010, p. 262). This can be visualized in terms of bottlenecks. In the example from Figure 1, the number of required separations for “1” and “7” (1 cut) differs from “1” and “4” (2 cuts). A larger cut set implies the network has to be dismantled more substantially to separate pairs, implying more similarity between nodes. In the Wikipedia network, we consider linkages between country pairs, which can take hundreds of cuts to separate. The statistic interprets similarity according to the commonality of topics and articles between them. The approach is slightly more complicated because hyperlinks run in a specific direction; a path along the network from one page to another is not necessarily reversible. Therefore, we measure how strongly nodes are independently connected through a connectivity min-cut statistic normalized by the average number of cuts needed to disconnect two nodes in the network.

Regression results

We test the hypothesis that longer ancestral separation correlates with less similar culture, with the novel culture variable derived by comparing the connectivity of topics and articles between countries’ Wikipedia pages. Descriptive statistics for the OLS regimes are listed in Table 3 below. All results consider treatments unique to network analysis for unobserved heterogeneity across neighboring observational units by using arbitrary clustering structures (Colella, et al., 2019). The regression takes the following form

$$TV_{i,j} = \beta_0 + \beta_1 N_{i,j} + \beta_k X_{i,j} + \varepsilon$$

where subscripts i,j refer to country-direction pair, TV refers to the different trade volume measures between countries, N refers to the connectivity statistic between pages, X is a vector of k controls between countries, and ε an error term. Because ancestry similarity owes much to geographic proximity, and to mitigate concern that the relationships merely go through geographic distance, we include controls for contiguity, geography, and water relationships.

The interesting feature of these data stems from separating culture into different categories. In Tables 2, the network was coarsened to only include pages from a relevant category tag along the measures discussed in the previous section. We consider eight dimensions from Table 1: business, education, government, language, military, religion, social, and technology. Across all regressions, we see a robust statistical relationship with expected signs. Given the construction of the connectivity variable, a larger coefficient implies a larger impact of culture on trade.

Table 2: Results of network cultural statistics regressed onto current foreign direct investment between countries.

	(1) business	(2) education	(3) government	(4) language	(5) military	(6) religion	(7) social	(8) technology
Network Connectivity	839.0*** (304.5)	1,428*** (490.5)	365.5** (143.9)	636.0*** (232.3)	744.8** (326.5)	748.2*** (288.9)	310.6*** (112.7)	908.8*** (316.7)
Contiguity controls	y	y	y	y	y	y	y	y
Geography controls	y	y	y	y	y	y	y	y
Water controls	y	y	y	y	y	y	y	y
Constant	-24,897** (9968.0)	-21,088** (8574.0)	-25,635** (10508.0)	-20,860** (8507.0)	-22,077** (9898.0)	-18,410** (7987.0)	-21,870** (9018.0)	-22,587** (9225.0)
Observations	1,763	1,763	1,763	1,763	1,763	1,763	1,763	1,763
Centered R-squared	0.173	0.167	0.171	0.144	0.158	0.148	0.153	0.168

*Robust standard errors in parentheses; *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.*

Table 3: Results of network cultural statistics regressed onto current net factor income from abroad between countries.

	(1) business	(2) education	(3) government	(4) language	(5) military	(6) religion	(7) social	(8) technology
Network Connectivity	31.83* (18.07)	62.82** (28.26)	14.13* (8.23)	32.40** (13.97)	23.2 (18.45)	31.92* (16.60)	13.13** (6.60)	33.25* (17.08)
Contiguity controls	y	y	y	y	y	y	y	y
Geography controls	y	y	y	y	y	y	y	y
Water controls	y	y	y	y	y	y	y	y
Constant	-288.8 (423.60)	-242 (359.70)	-327.3 (450.20)	-307.1 (346.00)	-105.2 (434.20)	-112 (339.60)	-256.1 (368.70)	-185.6 (357.70)
Observations	1,454	1,454	1,454	1,454	1,454	1,454	1,454	1,454
Centered R-squared	0.159	0.161	0.159	0.159	0.153	0.156	0.157	0.157

*Robust standard errors in parentheses; *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.*

Works Cited

- Abramitzky, R. (2015). 'Economics and the modern economic historian.' *The Journal of Economic History*, 75(4), 1240–1251.
- Acemoglu, D.; Robinson, J. A. (2012). *Why Nations Fail: The Origins of Power, Prosperity, and Poverty*. New York, NY: Crown Books.
- Alesina, A.; Devleeschauwer, A.; Easterly, W.; Kurlat, S.; Wacziarg, R. (2003). 'Fractionalization.' *Journal of Economic Growth*, 8, 55–194.
- Alesina, A.; Giuliano, P. (2015). 'Culture and institutions.' *Journal of Economic Literature*, 53(4), 898-944.
- Besley, T., & Persson, T. (2019). 'Democratic values and institutions.' *American Economic Review: Insights*, 1(1), 59-76.
- Blaydes, L., Grimmer, J., McQueen, A. (2018). 'Mirrors for princes and sultans: Advice on the art of governance in the medieval Christian and Islamic worlds.' *Journal of Politics*, 80(4), 1150–1167
- Chang, H. (2010). 'Institutions and economic development: Theory, policy, and history.' *Journal of Institutional Economics*, 7(4), 473–498.
- Colella, F.; Lalive, R.; Sakalli, S. O.; Thoenig, M. (2019) 'Inference with arbitrary clustering.' *IZA Discussion Paper* n. 12584
- Duhaime, E. P. (2015). 'Is the call to prayer a call to cooperate? A Field Experiment on the Impact of Religious Salience on Prosocial Behavior.' *Judgment and Decision Making*, 10(6), 593–596.
- Everett, J.; Haque, O.S.; Rand, D. (2016). 'How good is the Samaritan, and why? An experimental investigation of the extent and nature of religious prosociality using economic games.' *Social Psychological and Personality Science*, 7(3), 248–255.
- Giles, J. (2005). 'Internet encyclopaedias go head to head.' *Nature*, 438, 900–901.
- Gorodnichenko, Y.; Roland, G. (2011). 'Which dimensions of culture matter for long-run growth?' *American Economic Review*, 101(3), 429-98.
- Grajzl, P.; Murrell, P. (2019). 'Toward understanding 17th century english culture: A structural topic model of Francis Bacon's ideas.' *Journal of Comparative Economics*, 47(1), 111-135.
- Greenstein, S.; Feng, Z. (2012). 'Collective intelligence and neutral point of view: The case of Wikipedia.' *NBER*, Working Paper 18167.
- Guiso, L.; Sapienza, P.; Zingales, L. (2006). 'Does culture affect economic outcomes?' *Journal of Economic Perspectives*, 20(2), 23–48.
- Hayek, F.A. (2011 [1960]). *The Constitution of Liberty: The Definitive Edition*. Chicago, IL: University of Chicago Press.
- Henrich, J. (2000). 'Does culture matter in economic behavior? Ultimatum game bargaining among the Machiguenga of the Peruvian Amazon.' *American Economic Review*, 90(4), 973-979.
- Henrich, J. (2020). *The WEIRD People in the World*. New York, NY: FSG Publishing.
- Hodgson, G.M. (2006) 'What are institutions?' *Journal of Economic Issues*, 40(1), 1–25.
- Langlois, R. N.; Garzarelli, G. (2008). 'Of hackers and hairdressers: Modularity and the organizational economics of open-source collaboration.' *Industry and Innovation*, 15(2), 125–143.
- Mesgari, M.; Okoli, C.; Mehdi, M.; Nielsen, F.; Lanamki, A. (2015). 'The sum of all human knowledge: A systematic review of scholarly research on the content of Wikipedia.' *Journal of the Association for Information Science and Technology*, 66(2), 219-265.
- Meyer, R. (2013). '90% of Wikipedia's editors are male—Here's what they're doing about it.' *The Atlantic*, October 2013.
- Michalopoulos, S.; Papaioannou, E. (2016). 'The long-run effects of the scramble for Africa.' *American Economic Review*, 106(7), 1802-48.
- Michalopoulos, S.; Xue, M. (2021). 'Folklore.' *Quarterly Journal of Economics*, 136(4), 1993-2046.
- Newman, M. (2010). *Networks: An Introduction*. Oxford, UK: Oxford University Press.
- Norenzayan, A.; Shariff, A. (2008). 'The origin and evolution of religious prosociality.' *Science*, 322(5898), 58-62.
- North, D. (1990). *Institutions, Institutional Change, and Economic Performance*. Cambridge, UK: Cambridge University Press.
- Nunn, N. (2020a). 'History as evolution.' *NBER Working Paper* No w27706.
- Nunn, N. (2020b). 'The historical roots of economic development.' *Science*, 367, 1441.
- Petiška, E.; Moldan, B. (2019). 'Indicator of quality for environmental articles on Wikipedia at the higher education level.' *Journal of Information Science*, 47(2).
- Rubin, J. (2017). *Rulers, Religion, and Riches: Why the West Got Rich and the Middle East Did Not*. Cambridge, UK: Cambridge University Press.

- Spolaore E.; Wacziarg, R. (2016). 'Ancestry, language and culture.' In Ginsburgh V.; Weber S. (eds) *The Palgrave Handbook of Economics and Language*. New York, NY: Palgrave Macmillan.
- Spranz, R.; Lenger, A.; Goldschmidt, N. (2012). 'The relation between institutional and cultural factors in economic development: The case of Indonesia.' *Journal of Institutional Economics*, 8(4), 459–488.
- Steinsson, S. (2023). 'Rule ambiguity, institutional clashes, and population loss: How Wikipedia became the last good place on the internet.' *American Political Science Review*, 1-17.
- Viégas, F.B.; Wattenberg, M.; Dave, K. (2005). 'Studying cooperation and conflict between authors with history flow visualizations.' *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 4, 575–582.
- Wikistats. (2023). 'Wikimedia statistics 2.' *Wikimedia Foundation, Inc.* <https://stats.wikimedia.org/>. Accessed 22 September 2023.