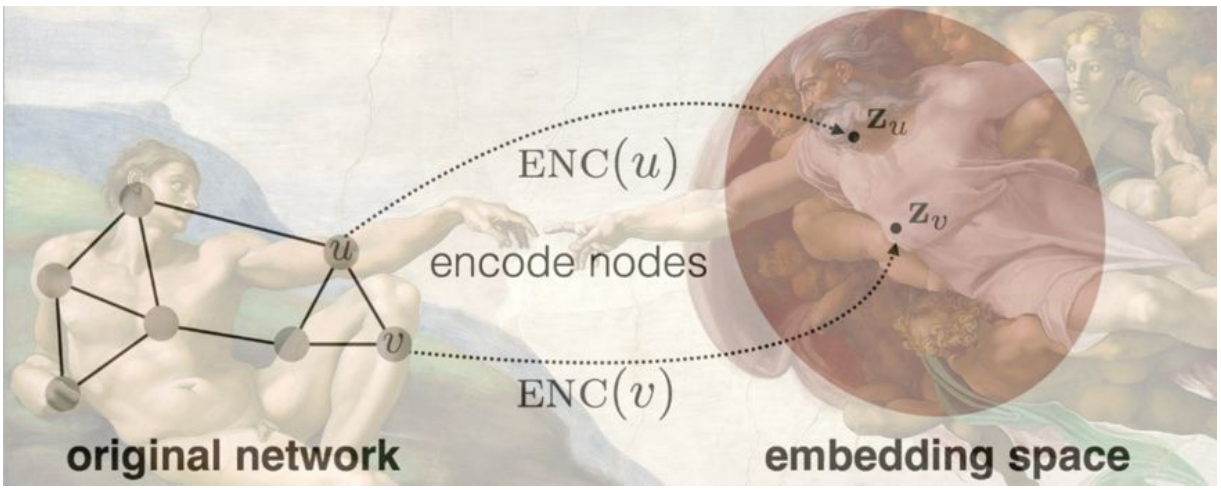
University of Pittsburgh
School of Computing and Information

INFSCI 3350 DOCTORAL SEMINAR: Network Embeddings to Model Science, Technology, and Economic Growth

Instructor: Lingfei Wu https://lingfeiwu.github.io/
Class No. 21891 (1050)
Days and Time: Tuesday, 12:00 pm - 2:50 pm
Room: IS 828



I. Course Description and Objectives

Network embedding, or Graph Representation Learning, is a task of learning vector representation of nodes on networks so that we can calculate these vectors to retrieve the network similarity between nodes easily. Node similarity can be defined in many ways, including the first-order similarity (nodes are linked), the second-order similarity (nodes share neighbors), or high-order associations (e.g., nodes occupy the same level in the "hierarchy" of networks). From this perspective, frequently used network embeddings such as "deepwalk" (Perozzi et al. 2014) or node2vec (Grover and Leskovec 2016) are only different in the definitions of node similarities, or, in natural language processing terms, strategies in retrieving the context nodes for a target node. The parameterizing and training of these models are the same (check Jure Leskovec's lecture for more details). For this reason, our introductory seminar focuses on simple node similarities (the first and second orders) and swallow embeddings (one hidden layer) before extending to complicated similarities and deep embeddings such as TransE (Bordes et al. 2013) or Graph Neural Networks (Kipf et al., 2017). Using simple yet scalable models sets us free from overfitting and increases the likelihood of discovering universal patterns across the complex social contexts.

Network embeddings emerge from the conjunction of complex networks (what to embed), manifold learning (how to embed), and neural networks (how to embed efficiently). We will discover how network embeddings are related to the important literature from these fields with a focus on the methodological advancement of network embeddings in *1) permitting simple models of diffusion by capturing the hidden geometry of networks; 2) revealing the common bases of node functions through their distributional representation; 3) recommending the optimal paths in networks by modeling the dependencies between nodes and links.* We will discuss how to leverage these reasoning and predictive power of network embeddings in understanding the coordination of knowledge and skills within and between economic agents ranging from workers, teams, companies, and cities to countries.

We use a "bulb string lights" metaphor to understand how talent is packed and coordinated in social systems. The limited cognition capacity of individual knowledge workers determines the limited progress of science, education, and economic growth that can be made at a time, reducing the complexity of social imagination down to its realization (Hidalgo 2015). Teams and labor divisions are invented to go beyond individual capacity by channeling individual wisdom into knowledge and skill networks such as companies, cities, and countries. However, this massive use and production of knowledge is not perfect - *not all good things scale up with population and connection*. As "string lights" networks evolve to be larger and more resilient, the individual "bulbs" are getting dimmer and more replaceable - the fragmentation of scholarship and the reskilling of workers. For example, the larger teams may generate more ideas (Wuchty, Jones, and Uzzi 2007), but not necessarily more innovative ideas (Wu, Wang, and Evans 2019). The slow diffusion of "know-how" (Hausmann et al. 2014) across experts who are trained to deliver specific tasks has become both the cause and consequence of the unscalability of integrated knowledge, slowing down the advance in science and technology and the growth of economic growth.

Can artificial intelligence help solve this problem of collective wisdom? Rather than naively scaling up networks, multiplying connections, and copying existing solutions, can it facilitate our deep thinkings into unknown scientific frontiers and cutivate disruptive innovations? More specifically, can it help recognize the moving frontiers of science and technology? Could it help decompose complex skills and knowledge into "atoms" and translate between professions and scientific fields? Would it make it possible to identify optimal skill paths for workers, teams, and companies from complex and rapidly changing landscapes? The importance of these questions and their answers is echoed by emerging fields such as the Science of Team Science (https://www.inscits.org/), the Science of Science (Fortunato et al. 2018), The Future of Work (Frey and Osborne 2013) and beyond. This seminar is build around these questions and upon finishing the seminar, the students are expected to

a. Contribute to the discussions on how science and technology progress and how economies grow;

b. Recognize how new network-embedding papers contribute to modeling knowledge production;
c. Applying network embeddings techniques to study how knowledge and skills coordinate.

Prerequisites

Network analysis, data mining, machine learning or relevant courses. Basic understanding in linear algebra, probability and statistics, and artificial neural networks are required. Adequate skills in Python is recommended (preferably experienced in Pytorch).

II. Attendance

Attendance is mandatory and will be recorded. Arriving late and leaving early without permission will affect your grade. If you must be absent, please inform in advance.

III. Grading

Grades are based on four major activities including class participation (5%), reading notes (30%), in-class presentations (30%), final project (35%). Assignments are due as scheduled, and grades on late work will be decreased by 20% per day late. The following table shows the translation between raw scores and letter grades:

Score	100-98	97-94	93-90	89-88	87-84	83-80	79-78	77-84	73-70	69-60	<60
Grade	A+	A	A-	B+	B	B-	C+	C	C-	D	F

IV. Course Design

Course arrangements: Each week, the instructor and the students will meet for three hours. There will be three 50-minutes sections with two 10-minutes break in between. The activities in each section are listed in each week's schedule.

For each week, there are required readings including conference papers, journal articles, or online materials. The student is expected to submit ONE reading note to cover all the readings. The notes should be submitted to the Courseweb Discussion Board by Sundays 11:50pm before the following class. The reading notes will be presented and discussed in the class. The following perspectives are encouraged, including connections to one of the research problems that the student is working on, critiques on limitations of the proposed method, or extensions to the original paper for following studies.

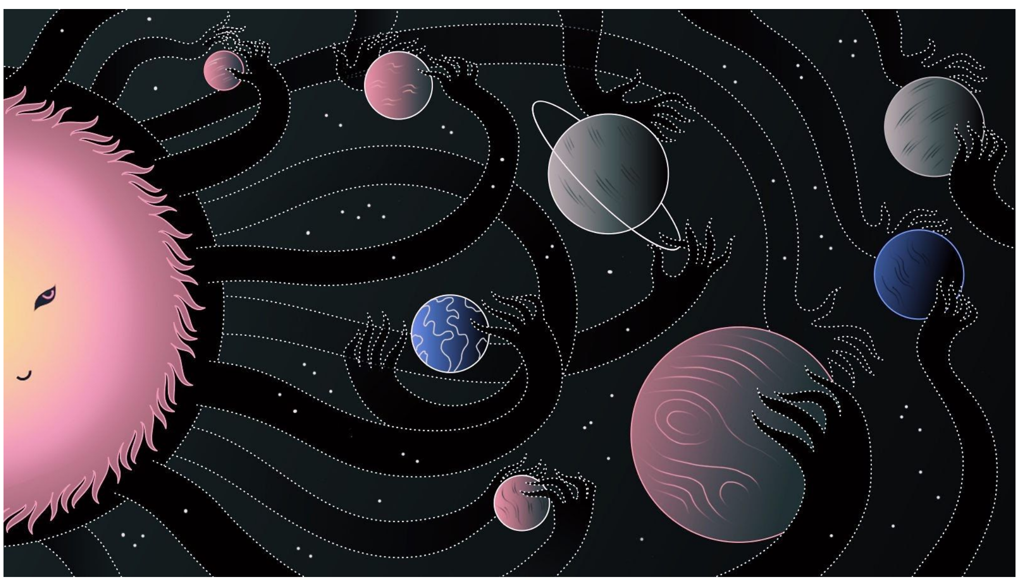
V. Term Projects

This is an individual, semester-long project. The goal is to design a research project that is related to the student's research interests. The project has to resolve some real-world problems using empirical data, and it should be research-oriented, rather than naively applying existing (e.g., GitHub) codes. The deliverable of the project should be a research article that includes an introduction and research question, related works, data and methods, and findings. It would be a bonus to the final project score if the paper is submitted for review at conferences or journals. Students are encouraged to synchronize this final project with other ongoing research projects including the PhD dissertation.

V. Course Organization.

This course includes three components organized by applied consequences, including 1) Identifying the moving frontiers of science and technology (weeks 1-5); 2) Discovering skill and knowledge atoms to translate between professions and scholarships (weeks 6-10); 3) Designing optimal skill paths for workers, teams, and companies (week 11-14). See the below parts for details.

1. Identifying the Moving Frontiers of Science and Technology (weeks 1-5)



Physics is governed by the duality between dynamics (Newtonian) and geometry (Einsteinian): complex dynamics on simple geometry equals simple dynamics on complex geometry. For example, the "preferential attachment" model of network growth (Barabasi and Albert 1999) is proven to equal near-neighbor linking on a hyperbolic space (Papadopoulos et al. 2012) in replicating the long-tail degree distribution, which is universally observed across real-world networks. In general, equations describing hidden geometry underlying networks can be translated into equations describing diffusion on networks, as demonstrated by (Brockmann and Helbing 2013) using disease diffusion between cities. This explains why word2vec, a geometry model of words, successfully predicts the diffusion of scholars' collective attention in the knowledge space towards new chemical materials (Tshitoyan et al. 2019).

Motivated by how Einstein discovered the geometrical nature of gravity as the spacetime curvature, we propose the "*geometry of ideas*" as an alternative to the current, social-network based models of the diffusion of information (Goel et al. 2016; Bakshy et al. 2012) and memes (Leskovec, Backstrom, and Kleinberg 2009; Kuhn, Perc, and Helbing 2014). Following the practice of calculating the "geodesic distance" between nodes using local linkages constructed in high-dimensional feature spaces (Tenenbaum, de Silva, and Langford 2000), the "geodesic distance" between two ideas is defined as the conditional probability of arriving one after the visit to the other. For example, a network can be constructed by extracting keywords as nodes from scientific literature with the weights on links equal to the frequency of co-occurrence. After this network is embedded, the dot product between node vectors should approximate their "geodesic distance" or the speed of diffusion from one to another, which characterizes the rate of "idea substitution", i.e., how fast the field progresses forward and scholars are aging and fading.

- **Week 1.** Brockmann, D., & Helbing, D. (2013). The hidden geometry of complex, network-driven contagion phenomena. *science*, 342(6164), 1337-1342.

  This paper proposes that commuting distance, defined as the logged inverse of the number of flight passengers between two cities, is a better measure than the geographical distance in predicting the spread of disease.

- **Week 2.** Tenenbaum, J. B., De Silva, V., & Langford, J. C. (2000). A global geometric framework for nonlinear dimensionality reduction. *science*, *290*(5500), 2319-2323.

  This paper proposes to extract the "geodesic distance", or "characteristic distance", between nodes from high-dimensional Euclidean feature spaces by calculating the similarity distance to link nodes and then only keep local linkages.

- **Week 3.** Papadopoulos, F., Kitsak, M., Serrano, M. Á., Boguná, M., & Krioukov, D. (2012). Popularity versus similarity in growing networks. *Nature*, *489*(7417), 537-540.

  The authors propose that near-neighbor linking on a hyperbolic space replicates the long-tail degree distribution and high clustering coefficient of real-world networks, thus their model is equal to the "preferential attachment" (Barabasi 1999) and "small-world" (watts and Strogatz 1998) models combined.

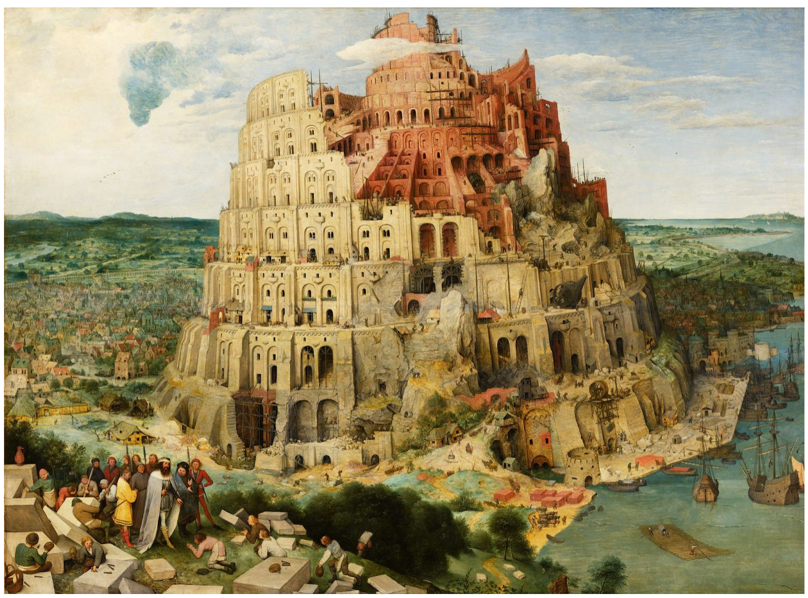
- **Week 4.** Tshitoyan, V., Dagdelen, J., Weston, L., Dunn, A., Rong, Z., Kononova, O., ... & Jain, A. (2019). Unsupervised word embeddings capture latent knowledge from materials science literature. *Nature*, 571(7763), 95-98.

The authors propose that word2vec (Mikolve 2013), a neural network linguistic model, successfully predicts the diffusion of scholars' collective attention in the knowledge space towards new chemical materials.

- **Week 5.** Levy, O., & Goldberg, Y. (2014). Neural word embedding as implicit matrix factorization. In Advances in neural information processing systems (pp. 2177-2185).

  This paper proves mathematically that term-context embedding (T_i*C_j) in word2vec skip-gram negative sampling (SGNS) implicitly models pointwise mutual information (PMI) between two words.

2. Discovering Skill and Knowledge Atoms to Translate between Professions and Scholarships (week 6-10)



Word2vec is based on the distributional hypothesis in linguistics (Harris 1964), i.e., the meanings of words are determined by the surrounding words, similar to the fact the functions of species are determined by the ecological context they live in (Jurafsky and Martin 2015). This philosophy connects word2ec to sparse PCA, which is a representation learning method that aims at finding a sparse representation of the input data in the form of a linear combination of basic, densely compressed vectors (Arora et al. 2018). In other words, word embeddings decompose meaning in a similar manner as how Newton's prism experiment splits lights or how the Fourier transform decomposes waves.

In analyzing knowledge and skill coordination, we can learn the vector representation of the keywords of knowledge or skills from scholarly or employment documents. We then can use PCA-like statistical methods to identify the orthogonal dimensions or "skill atoms" as the building blocks of complex human knowledge or capital. This permits the decomposition of complex tasks into simple functions and the translation of working experience between professions or scientific areas - a frequently encountered problem of deskilled workers or aging scholars is that they are not at a position to identify their own expertise under different names across the boundaries of industries or disciplines. We propose to use neutral networks for machine translation, but between the jargons of professions and scholarships rather than languages. This lays the ground for the ambitious plans of creating "GitHub of Science" or "Github of Work" which transparently manage different "versions" of science or economies through automatically extracting and documenting knowledge and skill combinations from papers and job ads.

- **Week 6.** Merton, R. K. (1961). Singletons and multiples in scientific discovery: A chapter in the sociology of science. *Proceedings of the American Philosophical Society*, *105*(5), 470-486.

  This paper discusses the pattern of "multiple discoveries", in which several scholars claim the same discovery independently. The author analyzed 264 multiple scientific discoveries and found them spread out across fields but concentrated in time.

  Simonton, D. K. (1978). Independent discovery in science and technology: A closer look at the Poisson distribution. *Social Studies of Science*, *8*(4), 521-532.

  The author revisits the Poisson model suggested by Price for multiple discoveries. The analysis of the empirical parameters suggests that science is the consequence of great times (social deterministic theory) rather than unique minds (great genius theory).

  Hill, R., & Stein, C. (2019). Scooped! Estimating Rewards for Priority in Science.

  The authors examine the impact of losing a priority race (colloquially known as getting"scooped") on subsequent publication and career outcomes and find that race winners receive more attention than losers, but that these contests are not winner-take-all - the effect is less stronger than what perceived subjectively as reveal by surveys.

- **Week 7.** Vilhena, D. A., Foster, J. G., Rosvall, M., West, J. D., Evans, J., & Bergstrom, C. T. (2014). Finding cultural holes: How structure and culture diverge in networks of scholarly communication. Sociological Science, 1, 221.

  The author analyzes the citation network and finds that communicative efficiency decays with citation distance in a field-specific way. These decay rates reveal hidden patterns of cohesion and fragmentation - the tendency to specialized jargon.

  Nissen, S. B., Magidson, T., Gross, K., & Bergstrom, C. T. (2016). Publication bias and the canonization of false facts. *Elife*, *5*, e21451.

  The authors model the community's confidence in a claim as a Markov process with successive published results shifting the degree of belief.

- **Week 8.** Arora, S., Li, Y., Liang, Y., Ma, T., & Risteski, A. (2018). Linear algebraic structure of word senses, with applications to polysemy. Transactions of the Association for Computational Linguistics, 6, 483-495.

  This paper shows that multiple word senses reside in linear superposition within word embeddings and that simple sparse coding can recover the vectors of these senses. The author extracted 2000 word senses and called them "discourse atoms".

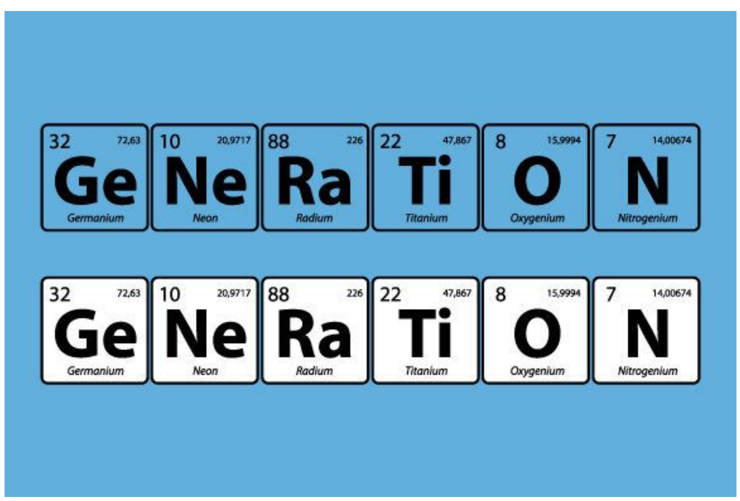
- **Week 9.** Arora, S., Li, Y., Liang, Y., Ma, T., & Risteski, A. (2018). Linear algebraic structure of word senses, with applications to polysemy. Transactions of the Association for Computational Linguistics, 6, 483-495.

  This paper shows that multiple word senses reside in linear superposition within word embeddings and that simple sparse coding can recover the vectors of these senses. The author extracted 2000 word senses and called them "discourse atoms".

- **Week 10.** Bordes, A., Usunier, N., Garcia-Duran, A., Weston, J., & Yakhnenko, O. (2013). Translating embeddings for modeling multi-relational data. In Advances in neural information processing systems (pp. 2787-2795).

  This paper introduces the model of "TranseE", which embeds both entities and their relations in the same Euclidean space. This allows for reasoning across the paths of knowledge graphs by multiplying corresponding vectors. "TranseE" can be viewed as a supervised version of word2vec.

3. Designing Optimal Skills Paths for Workers, Teams, and Companies (weeks 11-14)



An often overlooked fact is that word2vec produces two sets of embeddings - each node is encoded into both term embedding and context embedding. In the Skip-gram Negative Sampling (SGNS, one-to-many) architecture, IN and OUT matrices correspond to term and context embeddings, respectively. In the CBOW (many-to-one) framework, this is reversed.

It is proved that the term-context vector dot product implicitly models the pointwise mutual information (PMI) between items (Levy and Goldberg 2014), which quantifies to what extent two items co-occur more likely than random and has been used to identify collocation in linguistics (Rapp 2002) and complementary in economics (Neffke 2019). Meanwhile, the term-term vector dot product models the substitution between items (i.e., how similar or exchangeable two items are). Based on these observations, one can train embeddings of skills from large-scale literature of science, technology, and employment and model how workers, teams, and companies pack complement or substitute skills together to maximize value.

- **Week 10.** Rapp, R. (2002, August). The computation of word associations: comparing syntagmatic and paradigmatic approaches. In Proceedings of the 19th international conference on Computational linguistics-Volume 1 (pp. 1-7). Association for Computational Linguistics.

  This paper reviews the idea of the 1st and 2nd order associations between words, which are also called "syntagmatic" and "paradigmatic" relations, respectively, following Ferdinand de Saussure (de Saussure 2011). This paper also proposes to measure the 1st order association by co-occurrence and quantify the 2nd order association by comparing context word similarity.

  Levy, O., Goldberg, Y., & Dagan, I. (2015). Improving distributional similarity with lessons learned from word embeddings. Transactions of the Association for Computational Linguistics, 3, 211-225.

  This paper proposes that the term-term vector dot product in the SGNS models the 2nd order association, whereas the term-context vector dot product model the 1st order association. The authors suggest that combining these two vectors improves the performance of word2vec on many NLP tasks.

- **Week 11.** Woolley, A. W., Chabris, C. F., Pentland, A., Hashmi, N., & Malone, T. W. (2010). Evidence for a collective intelligence factor in the performance of human groups. science, 330(6004), 686-688.

  Psychologists have repeatedly shown that a single statistical factor (IQ) emerges from the correlations among people’s performance on a wide variety of cognitive tasks. Using the similar approach, this paper finds that (GQ) exists, explaining a group’s collective intelligence which is correlated with the average social sensitivity of group members and the equality in distribution of conversational turn-taking, irrelevant to the IQ of group members.

  Wu, L., Wang, D., & Evans, J. A. (2019). Large teams develop and small teams disrupt science and technology. Nature, 566(7744), 378-382.

  This paper suggests that small and large teams are different in nature. Small teams ask questions and disrupt existing theories. Large teams answer questions and stabilize established paradigms.

- **Week 12.** Nalisnick, E., Mitra, B., Craswell, N., & Caruana, R. (2016, April). Improving document ranking with dual word embeddings. In Proceedings of the 25th International Conference Companion on World Wide Web (pp. 83-84).

  This paper analyzes large-scale text data and shows how term-term and term-context embeddings model the 2nd and 1st order associations, respectively.

  Sauer, C., Haigh, A., & Rachleff, J. Cooking up Food Embeddings.

  This paper trains food ingredient embeddings by modeling complement food ingredient pairs using the term-context vector dot product and substitute pairs using the term-term vector dot product.

- **Week 13.** Boudreau, K. J., Lacetera, N., & Lakhani, K. R. (2011). Incentives and problem uncertainty in innovation contests: An empirical analysis. Management science, 57(5), 843-863.

  This paper analyzes programming contest data from Topcoder and suggests that for uncertain tasks many small teams are better, whereas for tasks high in certainty a few large teams is the best.

  Mason, W., & Watts, D. J. (2012). Collaborative learning in networks. Proceedings of the National Academy of Sciences, 109(3), 764-769.

  This paper designs online games and finds that collective exploration improved average success over independent exploration as good solutions could diffuse through the network. Also efficient networks outperform inefficient networks.

- **Week 14.** Hidalgo, C. A., Klinger, B., Barabási, A. L., & Hausmann, R. (2007). The product space conditions the development of nations. Science, 317(5837), 482-487.

  This paper shows that countries are moving slowly on the "sticky" landscape of tech know-how.

  Neffke, F. M. (2019). The value of complementary co-workers. Science Advances, 5(12), eaax3370.

  This paper constructs two education degree networks (the first one linked complement ingredients as they are co-used in economic establishments, the second one linked substitutive ingredients as they prepare students for the same professions). The authors found that having co-workers of degrees complementing one's own permits a salary premium.

  Alabdulkareem, A., Frank, M. R., Sun, L., AlShebli, B., Hidalgo, C., & Rahwan, I. (2018). Unpacking the polarization of workplace skills. Science advances, 4(7), eaao6030.

  This paper shows the polarization in skill networks between cognitive and physical skills and reveals the projection of this polarization at the level of professions and cities and the consequential constraints to career mobility.

References

Arora, Sanjeev, Yuanzhi Li, Yingyu Liang, Tengyu Ma, and Andrej Risteski. 2018. "Linear Algebraic Structure of Word Senses, with Applications to Polysemy." *Transactions of the Association for Computational Linguistics* 6 (December): 483–95.

Bakshy, Eytan, Itamar Rosenn, Cameron Marlow, and Lada Adamic. 2012. "The Role of Social Networks in Information Diffusion." In *Proceedings of the 21st International Conference on World Wide Web*, 519–28. WWW '12. New York, NY, USA: Association for Computing Machinery.

Barabasi, A. L., and R. Albert. 1999. "Emergence of Scaling in Random Networks." *Science* 286 (5439): 509–12.

Bettencourt, Luís M. A., José Lobo, Dirk Helbing, Christian Kühnert, and Geoffrey B. West. 2007. "Growth, Innovation, Scaling, and the Pace of Life in Cities." *Proceedings of the National Academy of Sciences of the United States of America* 104 (17): 7301–6.

Brockmann, Dirk, and Dirk Helbing. 2013. "The Hidden Geometry of Complex, Network-Driven Contagion Phenomena." *Science* 342 (6164): 1337–42.

Fortunato, Santo, Carl T. Bergstrom, Katy Börner, James A. Evans, Dirk Helbing, Staša Milojević, Alexander M. Petersen, et al. 2018. "Science of Science." *Science* 359 (6379). https://doi.org/10.1126/science.aao0185.

Frey, Carl Benedikt, and Michael Osborne. 2013. "The Future of Employment." http://sep4u.gr/wp-content/uploads/The_Future_of_Employment_ox_2013.pdf.

Goel, Sharad, Ashton Anderson, Jake Hofman, and Duncan J. Watts. 2016. "The Structural Virality of Online Diffusion." *Management Science* 62 (1): 180–96.

Harris, Zelig S. 1964. "Distributional Structure. WORD, 10: 146--162. Reprinted in J. Fodor and J. Katz, The Structure of Language: Readings in the Philosophy of Language." Prentice-Hall.

Hausmann, Ricardo, César A. Hidalgo, Sebastián Bustos, Michele Coscia, Alexander Simoes, and Muhammed A. Yildirim. 2014. *The Atlas of Economic Complexity: Mapping Paths to Prosperity*. MIT Press.

Hidalgo, Cesar. 2015. *Why Information Grows: The Evolution of Order, from Atoms to Economies*. Basic Books.

Jurafsky, Daniel, and James H. Martin. 2015. "Speech and Language Processing, Chapter Chapter 19: Vector Semantics." *Prentice Hall, 3rd Edition. Draft of August* 24: 2015.

Kuhn, Tobias, Matjaž Perc, and Dirk Helbing. 2014. "Inheritance Patterns in Citation Networks Reveal Scientific Memes." *Physical Review X* 4 (4): 041036.

Leskovec, Jure, Lars Backstrom, and Jon Kleinberg. 2009. "Meme-Tracking and the Dynamics of the News Cycle." In *Proceedings of the 15th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 497–506. KDD '09. New York, NY, USA: Association for Computing Machinery.

Levy, Omer, and Yoav Goldberg. 2014. "Neural Word Embedding as Implicit Matrix Factorization." In *Advances in Neural Information Processing Systems 27*, edited by Z. Ghahramani, M. Welling, C. Cortes, N. D. Lawrence, and K. Q. Weinberger, 2177–85. Curran Associates, Inc.

Neffke, Frank M. H. 2019. "The Value of Complementary Co-Workers." *Science Advances* 5 (12): eaax3370.

Papadopoulos, Fragkiskos, Maksim Kitsak, M. Ángeles Serrano, Marián Boguñá, and Dmitri Krioukov. 2012. "Popularity versus Similarity in Growing Networks." *Nature* 489 (7417): 537–40.

Rapp, Reinhard. 2002. "The Computation of Word Associations: Comparing Syntagmatic and Paradigmatic Approaches." In *COLING 2002: The 19th International Conference on Computational Linguistics*. https://www.aclweb.org/anthology/C02-1007.pdf.

Saussure, Ferdinand de. 2011. *Course in General Linguistics*. Columbia University Press.

Tenenbaum, J. B., V. de Silva, and J. C. Langford. 2000. "A Global Geometric Framework for Nonlinear Dimensionality Reduction." *Science* 290 (5500): 2319–23.

Tshitoyan, Vahe, John Dagdelen, Leigh Weston, Alexander Dunn, Ziqin Rong, Olga Kononova, Kristin A. Persson, Gerbrand Ceder, and Anubhav Jain. 2019. "Unsupervised Word Embeddings Capture Latent Knowledge from Materials Science Literature." *Nature* 571 (7763): 95–98.

Wimsatt, William C., and William Kurtz Wimsatt. 2007. *Re-Engineering Philosophy for Limited Beings: Piecewise Approximations to Reality*. Harvard University Press.

Wuchty, Stefan, Benjamin F. Jones, and Brian Uzzi. 2007. "The Increasing Dominance of Teams in Production of Knowledge." *Science* 316 (5827): 1036–39.

Wu, Lingfei, Dashun Wang, and James A. Evans. 2019. "Large Teams Develop and Small Teams Disrupt Science and Technology." *Nature* 566 (7744): 378–82.