MIT Connection Science
the technology of innovation

2/23/2018

Human Interaction, Idea Flows, and Wealth Generation

Alex Pentland, MIT
pentland@mit.edu

New sources of "big data" such as geo-located communications and purchase data allow measurement of human interaction patterns across entire economies at a fine granularity. Most obviously this offers the ability to measure foot traffic to stores, employee work patterns, truck deliveries, and similar patterns related to the economic health of companies. Less obviously it allows measurement of interaction between different communities, and discovery of patterns which predict rates of innovation and economic growth. We find that measures of diversity and volume of physical interaction between communities (clusters of interaction) are strong predictors of economic productivity and growth, thus providing both new investment opportunities and new methods of risk assessment. Access and use of these data raise privacy and security risks, and we describe how these challenges can be controlled.

The wide adoption of mobile phones, social media, digital payments and similar technologies has dramatically changed our ability to understand our neighborhoods, cities, and countries: now there are data about the world and about each of us everywhere. With data at this scale and granularity, you can get a picture of society that was unimaginable only a few years ago. This ability to see human behavior continuously and quantitatively has created a new science called computational social science [1], and it is beginning to transform disciplines such as sociology, political science, and psychology [2, 3].

These geo-located data are already helping to better understanding global patterns of human behavior and helping decision makers to tackle problems of societal importance, such as: monitoring socio-economic deprivation [11, 10, 6, 12] and crime levels [4, 13, 14], mapping the propagation of diseases [14, 15, 16, 9], and understanding the impact of natural disasters, environmental risks, and other emergencies [7, 8, 4, 5], etc. These successes have encouraged the UN to identify these new data methods as the "data revolution" that will allow countries to measure progress toward the Sustainable Development Goals (SDGs), including the goals of zero poverty, zero hunger, good health, clean water and energy, gender equality, sustainable cities, quality education, etc. by 2030 [6].

This same data revolution is also beginning to transform economics, finance, and investment. In this paper I will present a sample of the insights we can obtain by using these new geo-location data resources, using data from a wide variety of cities in the US, EU, and Asia, and also how we can safely use these new data resources to build better investment and risk management systems.

Wealth Comes from Finding New Opportunities

What do these new data and analytics tools show? The short answer is that we are less individuals (rational or not) than we are creatures of our social networks. For instance, Figure 1 shows data from a sample of 100,000 randomly selected people in a mid-income country and compares their ability to hear about new opportunities (measured by the diversity of the communities that they interact with) to their income.

This example demonstrates that people who have more open networks, and thus more access to new ideas and opportunities, make more money. Moreover, this is not just an artifact of the way we measured their access to opportunities, because you can get the same result looking at the diversity of the jobs of the people they interact with,

or the diversity of locations of the people that they interact with. Surprisingly, if you compare people who have less than a 7th grade education to the entire population, the curve moves only a little to the left. If you look at people with college education or more, the curve moves only a little bit to the right. The variation that has to do with education is small when compared with the variation that has to do with diversity of interaction.

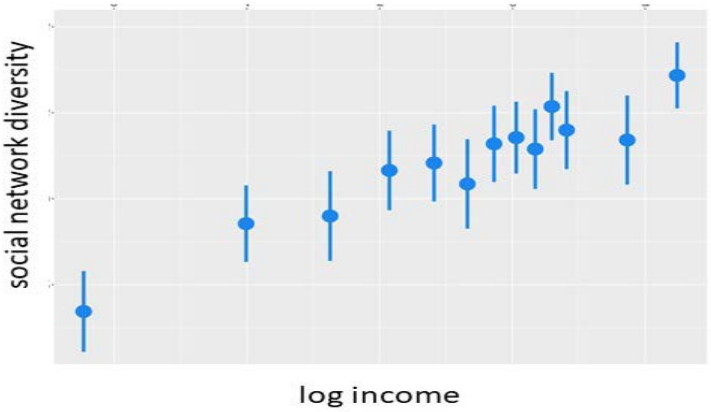


Figure 1. As people interact with more diverse communities, their income increases (100,000 randomly chosen people in mid-income country) [19].

You may wonder if greater network diversity *causes* greater income or whether it is the other way around. The answer is that greater network diversity indeed causes greater income on average…this is the idea of weak ties bringing new opportunities…but it is also true that greater income causes social networks to be more diverse.

The view that interaction patterns predict the flow of new ideas from one community to another, and that the flow of ideas accounts for a substantial portion of economic growth is also becoming established for interactions between nations [17, 18]. Consequently nations with unusually large volumes of interaction can, over the long run, be expected to have convergence in skills, technology, and productivity. What is surprising about the work presented here is the demonstration that the connection between interaction patterns, idea flow, and economic growth is also a major effect at the scale of individuals, neighborhoods, and cities. As a consequence observed patterns of interaction can be used to help "credit score" individuals, small businesses, real estate projects, and regional infrastructure projects.

In Western society, we generally assume that individual features far outweigh social network factors. However it may be more accurate to conceive of humans as a species who are on a continual search for new opportunities, and in which social networks serve as a major, and perhaps the greatest, resource for finding opportunities. Like all other social species, our lives consist of a balance between habits that allow us to exploit our environment and exploration to find new opportunities [20], as illustrated in Figure 2.

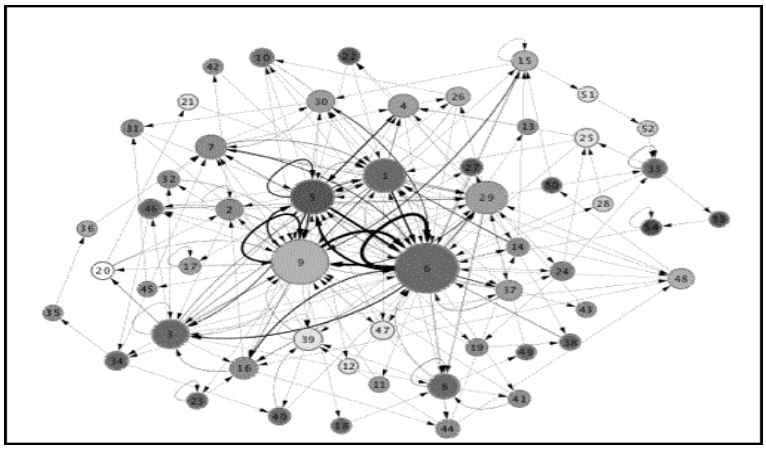


Figure 2: Foraging behavior is characteristic of human financial patterns around the world. This figure shows credit card purchase locations of a single person over one month. Circle size is proportional to purchase frequency, and arrows show the likelihood of transition between the purchase locations [20].

In the animal literature this is known as *foraging behavior*. It's the tension between exploitation and exploration, and it also the character of normal human life. When we examined credit card data for 100 million people in the US [20], we can see that people are immensely predictable. Most of people's purchases are habitual and occur at a small number of places (large circles) and the sequence of their purchases is highly predictable. For instance, if I know what you purchase in the morning I can know, with 90% plus odds of being right, what you will purchase in the evening. But every once in a while people break free of their daily habits and explore new places. They typically visit these new locations only very occasionally (small circles) and these exploratory sequences of purchase behavior are extremely unpredictable.

Moreover, when you find individuals who do not show this pattern, they are almost always sick or stressed in some way. You can tell whether a person's life is healthy in a general sense – both mental and physical – by whether they show this most basic biological rhythm or not [21]. This tendency is regular enough that one of the largest health services in the US is using this to keep track of the health of at-risk patients (see, for instance, our spin-off company ginger.io). Similarly, changes in a person's exploration pattern are often indicative of financial worries. In our experiments we find we can use the pattern of exploratory behavior to accurately predict whether an individual will have financial troubles with approximately three months advance warning [22].

If you turn this foraging behavior inside out, and look at the dual of all the individual purchase networks, you obtain the visitation network from the point of view of stores and other commercial venues. The pattern of a store patronage, e.g., how many people, from which demographics and how far they travel to get to the store, is quite informative of the store's future financial health. Figure 3 shows a recent result produced by the Thasos Group (one of my spin-off companies) and provided to investors by Bloomberg (see http://thasosgroup.com).

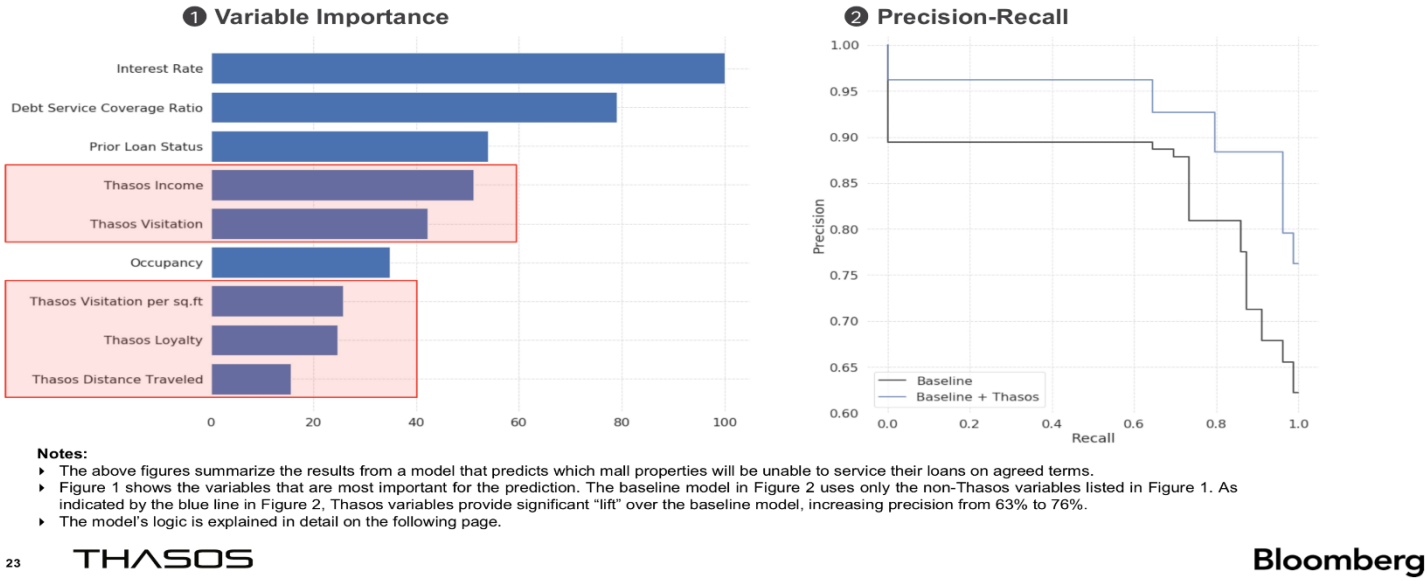


Figure 3. The pattern of visitation predicts financial viability of stores (http://thasosgroup.com).

Finally, if you combine this idea of foraging for novelty with the concept that diverse networks bring greater opportunities and greater income, you would expect that cities that facilitate connecting with a wide range of people would be wealthier. To test this hypothesis we gathered data from 150 cities in the US and 150 cities in the EU and examined the patterns of physical interactions between people [23].

We found that if a city's infrastructure facilitates more diverse physical interactions, then over the long term the citizens make more money. In Figure 5 below, you can see this model predicts GDP per square kilometer extremely accurately, in both the US and EU. What this suggests is that the factors that we usually think about -- investment, education, infrastructure, institutions -- may not be the direct cause of GDP growth. Instead they may make a difference primarily because they help or hinder the search for new opportunities. The main driver of progress in society may be the search for new opportunities rather than people's skills or capital investment.

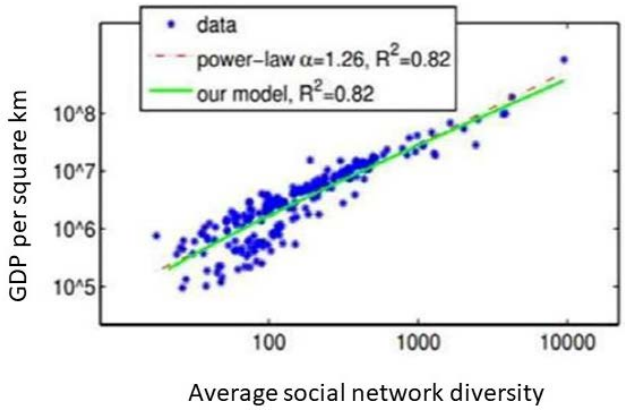


Figure 5. As face-to-face communication within a city allows interaction between more diverse communities, the city wealth increases. Data from 150 cities in the EU, pattern is similar to that found in the US [23].

This new computational social science understanding of human behavior and society in terms of networks, the search for new opportunities and the exchange of ideas might best be called Social Physics, a name coined two centuries ago by Auguste Comte, the creator of sociology. His concept was that the spread of ideas shaped the development of society in a regular and predictable manner. While his theories were in many ways too simplistic, the recent successes of computational social science show that he was going in the right direction. It is the *flow* of ideas and opportunities between people that drives society, providing quantitative predictions at scales ranging from small groups, to companies, cities, and even entire countries [24].

Building Social Bridges: opportunities for investment

Cities are a great example of how the process of foraging for new opportunities shapes human society. Cities are major production centers of society, and as we have already seen, cities in which it is easy to search for new opportunities are wealthier. Long term economic growth is primarily driven by innovation in the society, and cities facilitate the human interaction and idea exchange needed for good ideas and new opportunities to spread. For example, success on Wall Street often involves knowing about new events before the majority of other people. In these sorts of environment, the informational advantages of extreme spatial proximity become very high. This may explain why Wall Street remains in a tiny physical area in the tip of Manhattan. The spatial concentration of economic actors increases productivity at the firm level by increasing the flow of new ideas, both within and across firms.

Our evidence suggests that bringing together people from diverse communities will be the best way to construct a vibrant, wealthy city. When we examine flows of people in real cities we find that their physical infrastructure limits physical mixing more than we usually assume. People who live in one neighborhood typically work and shop in only a few other neighborhoods. Consequently, there are a small number of *social bridges* formed by the pattern of physical interaction between their home neighborhood and other neighborhoods. When neighborhoods are connected by social bridges, so that their residents spend time together, whether at work or at play, they learn from each other. Over time, connected neighborhoods begin to adopt similar behaviors and attitudes, along with similar skills, knowledge, and productivity.

Marketing. In fact, what we see is that all sorts of behaviors, such as what sort of clothes they buy, how they deal with credit cards, even behavioral diseases (like diabetes or alcoholism) flow mostly within groups connected by social bridges. Behavior patterns don't follow demographic categories nearly as closely. In a recent study of a large European city, for instance, we found that social bridge structure was more than *300% better* at predicting a wide range of behaviors than were demographics such as age, gender, income, and education (see Table 1). This means that investment systems based on a demographic stratification of the population (i.e., virtually all consumer finance systems) are performing far worse than they would if they were based on a behavior-based stratification [25].

Indicator	Pearson's r (part. corr.) with # co-visits	p-value computed using QAP
# Social Bridges	**0.7063**	<1e-4
Age	0.1327	<1e-4
Gender	0.1586	<1e-4
Marital Status	0.0834	1.4e-4
Education	0.1818	<1e-4
Working Style	0.0472	0.0093
Income	0.1775	<1e-4
Socio-demographics + Income	0.2261	<1e-4

Table 1. Social bridges, which are the pattern of physical interaction between neighborhoods, predicts purchasing and financial decisions three times better than traditional demographics [25].

Groups of neighborhoods joined by rich social bridges form local cultures. Consequently, by knowing a the places that a person hangs out in, you can tell a great deal about their preferences and behaviors. The process of learning from each other by spending time together means that ideas and behaviors tend to spread mostly within the cluster, but not further. A new type of shoe, a new genre of music, a political viewpoint, will spread within a cluster of neighborhoods joined by social bridges, but will tend not to go across cluster boundaries to other places. Marketers talk about influencers changing people's minds, but the more powerful effect is that when people spend time together they begin to mimic each other, they learn from each other, and they adopt similar behaviors and attitudes.

The social bridges idea is based on the observation that individuals can commonly be identified as part of a community with relatively homogeneous behavior based on where they live, work, and shop. Where you invest your most valuable resource – time -- reveals your preferences. Each community typically has access to different pools of information, opportunities, or offers different perspectives. Diverse interactions should then increase a population's access to the opportunities and ideas required for productive activity and economic growth.

Investment. When we apply this logic to neighborhoods in the US, Asia, and Europe, we find that diversity of social bridges (physical interaction) predicts 45% of the variance in year-on-year GDP growth, as illustrated in Figure 6. In other words, the effect on GDP of diversity of idea flow via the social bridges is about as important as the effect of population density. If we compare the explanatory strength of interaction diversity with other variables such as average age or percentage of residents who received a tertiary education, we find that these traditional measures are substantially weaker at explaining economic growth than social bridge diversity [26]. This means that models and systems that depend only on measurements of population, education, and financial investment may be missing the main effects.

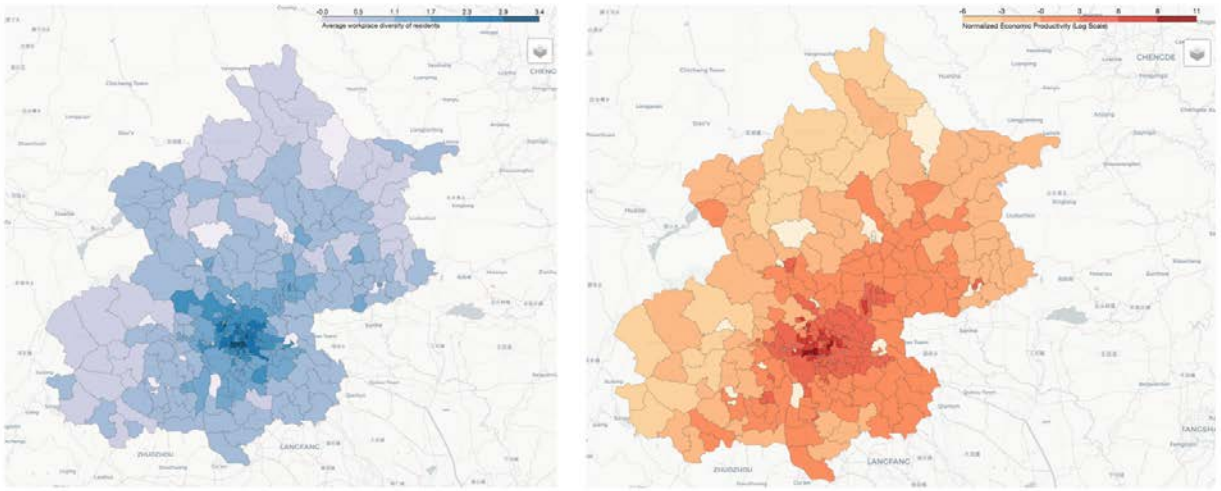


Figure 6. (left) Social bridges diversity (i.e., diversity of physical interactions between neighborhoods) and (right) year-on-year economic growth for neighborhoods within the city of Beijing. The diversity of social bridges predicts 45% of the variance in year-on-year economic growth; similar results are obtained for both US and EU cites [26].

A

Regression Results of Diversity and Innnovation

Variables	Intellectual Properties (Poisson)	Commercializing Performance (OLS)
Culture Diversity (CD)	2.581***(12.66)	1.214***(4.05)
Industries Diversity (ID)	1.167***(4.99)	1.902***(5.72)
Education Level Diversity (ED)	1.034***(5.32)	0.891***(3.94)
Employees (SS)	0.000***(5.10)	0.001***(6.10)
R&D Expenditures (RD)	0.007***(5.68)	0.012***(3.18)
Cont. (Cons)	2.956*** (43.31)	9.190***(94.55)
R^2	0.2620	0.2700
Samples	3255	3255

其中： *** 1% significant ， **5% significant， *10% significant

B

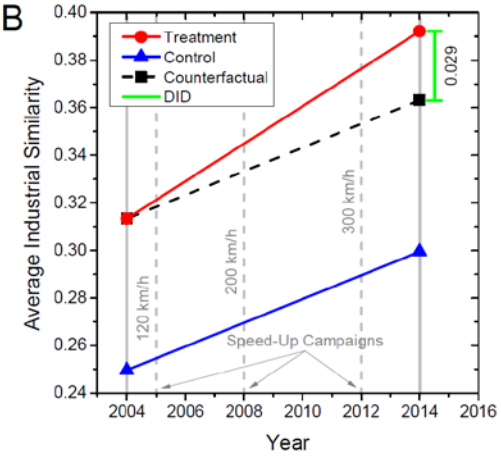


Figure 5. (A) a survey of all Chinese startups in government-sponsored incubators, showing that cultural diversity is the biggest factor in successful launch and funding of startups [27]. (B) Figure showing that facilitating travel between two cities (here by high-speed rail) is effective at spreading specialized commercial capacity (e.g., IT firms, or financial firms) from one city to the other [28]. In both cases mixing communities with different expertise and practice resulted in creation of successful companies that leveraged the skills of the different communities.

The view that physical mixing between different communities strongly influences the rate of innovation and financial growth is reinforced by two other studies recently completed in China. The first study looked at factors that were important in success of new businesses by performing a survey of startups launched from 3,255 government-sponsored incubators throughout China [27]. Figure 5(a) shows the result that cultural diversity was the biggest factor in successful launch and funding of startups, and that diversity of industrial experience was most important in the subsequent growth and scaling of these startup companies. The second study looked at whether investment in high-speed rail infrastructure promoted the spread of companies with specialized commercial expertise [28]. The blue solid line at the bottom in Figure 5(b) shows the increasing rate of company creation in all cities in China. The red solid line at the top of Figure 5(b) shows the rate of specialized company creation in cities that become connected to cities that had pre-existing companies with the same specialized expertise. In other words, when a city "X" with few information technology (IT) companies is connected by high speed rail to a city "Y" that already has many IT firms, the rail connection strongly promotes creation of more IT firms in city "X." The spread of ideas because of these new rail connections produced an increase of company creation within specialty or category that was almost 50% greater for connected cities than the general rate of company creation. In both studies mixing communities with different expertise and practice resulted in greater rates of company creation.

In his book "Moral Sentiments" Adam Smith said that it is the peer-to-peer exchanges of ideas and goods that determine norms and behaviors, and this is supported by our research. It is more accurate to think about society is in terms of interaction groups and idea flow between these groups rather than only in terms of the demographics of individuals. The idea of social bridges is often a more powerful concept than demographics, because social bridges are a major conduit by which people influence each other. By understanding these social bridges, we can begin to build better investment systems and create a more innovative society.

Optimizing Opportunity

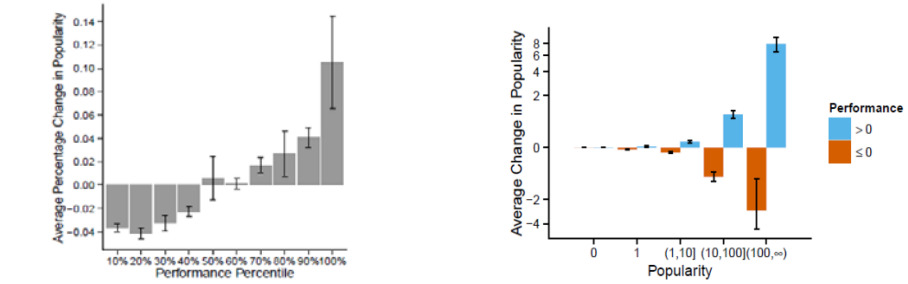
The above observations suggest that social structure that optimizes search for new ideas and of opportunities is a major factor in economic development. Assuming that exploration for new opportunities and ideas is central for development, then we should ask how humans trade off the cost of exploration for new opportunities versus

investment in familiar opportunities. To answer this question I will turn to data about financial investing which provides clear, simple and well-developed example of the trade-off between exploiting known opportunities and exploring for new ones.

To illustrate this trade-off in human decision making it is useful to look at Bayesian decision methods, specifically the class of algorithms known as Thompson sampling, which provide minimum investment regret while attempting to maximize return. These methods are used to choose among alternative actions when the potential for profit is unknown or uncertain. The core idea associated with these analysis methods is that when decision makers are faced with a wide range of alternative actions, each with unknown payoff, they must select actions to discover those that lead to the best payoffs and at the same time exploit the actions that are currently believed to be the best in order to remain competitive against opponents. This is the same idea as animals foraging for food, or people searching for new opportunities while still making a living.

In a social setting the payoff for each potential action can be easily and cheaply determined by observing the payoffs of other members of a decision maker's social network. This use of social learning dramatically improves both overall performance and reduces the cognitive load placed on the human participants. The ability to rapidly communicate and observe other decisions across the social network is one of the key aspects of optimal social learning and exploration for opportunity.

As an example, we recently examined how traders maximize their returns by the sharing of strategic information within a social-network stock-trading site where people can see the strategies that other people choose, discuss them, and copy them. We analyzed some 5.8 million transactions by day traders on a "social trading" platform called eToro, and found that the groups of traders who fared the best approximated the distributed Thompson sampling strategy.



a **b**

Figure 6: Social trading (a) Financial traders use recent performance in making decisions, but (b) this is modulated by using popularity signals as a prior probability that the trade is good, a strategy called distributed Thompson sampling. Using this strategy can produce much better ROI as compared to individual trading [29].

As shown in Figure 6, we found that traders used recent performance in making decisions, as expected, and exhibited an asymmetry due to loss aversion. However traders modulated decisions based on performance signals by using popularity signals as a prior probability that the trade will be good, a strategy called distributed Thompson sampling [29]. In fact, on the eToro platform the popularity of a trader (number of followers) is indeed a good estimate of the traders quality (average return on investment). Significantly, using this strategy results in much greater return on investment as compared to the strategies used by individuals trading alone.

Trustworthy, Clean Data

Today's "data ecology" is transforming due to the exponential growth of mobile and ubiquitous computing, together with the spread of big data analysis. Increased media coverage of cybersecurity breaches and intelligence scandals are having a dramatic impact on people's concerns about personal data sharing and security. The surge of mobile transactions, micropayments, and connected sensors in both private and public spaces is expected to further exacerbate this tension. We need a "new deal on data" where security concerns are matched with transparency, control and privacy, and are designed into the core of any data-driven service [30, 31].

Key to realizing trusted "next generation" data architectures is creating systems that seriously address the challenges of privacy, data ownership, data securitization, and cybersecurity. To address these issues, we have developed the *Open Algorithms* (OPAL) architecture, as well as high-security machine learning methods which build on OPAL.

The concept of OPAL is that instead of copying or sharing data, algorithms are sent to existing databases, executed behind existing firewalls, and only the encrypted results are shared. This minimizes opportunities to attack databases or divert data for unapproved use.

The Open Algorithms paradigm seeks to address the increasing need for individuals and organizations to share data in a secure, privacy-preserving manner. Today there are a number of open challenges with regards to the information sharing ecosystem:

- *Data is siloed*: This makes data unavailable to support good decision making.
- *Privacy is inadequately addressed*: European regulations and other forces are beginning to address this problem, but it is still far from solved.
- *Security is failing:* The current "firewall" architecture is fundamentally inadequate, as the almost daily reports of hacking events and lost customer data demonstrate.

The OPAL principles for secure, privacy-preserving sharing of insights are simple and relatively easy to implement. Through the MIT Trust Data Consortium (trust.mit.edu) we have created an alliance of countries and multinational and multilateral organizations that supporting the OPAL paradigm as the foundation of a sustainable digital ecology. Several of our Trust Data Consortium sponsors, in Israel, China, and France, are working with us to field national-scale "safe data" systems built on OPAL. One of our Trust Data Consortium sponsors, Helios Data, is working with us to commercialize OPAL into a system that ensures that data is "clean" from privacy, regulatory and third-party ownership constraints. Such clean data will allow companies to develop new streams of revenues from monetizing their valuable data assets.

Beyond the OPAL framework, it is now possible to perform machine learning and AI directly on encrypted data, without ever decrypting or otherwise exposing the data. Another of our spin-out companies, Endor.com, is already performing this surprising feat, using its "social physics" technology to serve many large multinational companies. Because the Endor.com technology works only with encrypted data, no data is ever shared, and machine learning occurs without revealing the queries to others so that the Endor system can discover value in other people's encrypted data assets without endangering commercial secrets, tripping regulatory restrictions, or violating privacy. The ability to do machine learning and test the value of data without ever exposing unencrypted data promises to revolutionize data-dominated financial systems.

Conclusions

The foundations of modern of Western society, and of rich societies everywhere, were laid in the 1700s in Scotland by Adam Smith, John Locke, and others. The understanding of ourselves that this early social science created is that humans are rational individuals and independent thinkers who are driven by self-interest. This viewpoint is built into

every part of our society now – we use markets, we take direct democracy as our ideal of government, and our schools focus on training students to have better analytics skills. As a consequence the rational individual model has become the bedrock of financial systems and strategies.

But this rational individual model is much too simple – and it is not just the rational "self-interested" part that is questionable, but more importantly the part about humans acting as individuals who are independent of everyone else. Our behavior is strongly influenced by those around us and our ability to thrive is largely due to learning from other people's experiences. We are not individuals but rather members of a social species, and the consequences for financial systems and investment strategies are immense, as I have tried to suggest in this paper.

The idea of "rational individuals" reached its current form when mathematicians in the 1800's tried to make sense of Adam Smith's observation that people "…are led by an invisible hand to … advance the interest of the society, and afford means to the multiplication of the species." These mathematicians found that they could make the invisible hand work if they used a very simplified model of human nature: people act only to benefit themselves (they are "rational"), and they act alone, independent of others (they are "individuals"). While this may not be a bad first approximation, it fails in the end because it is people influencing each other, peer-to-peer, that causes financial bubbles, cultural change, and it is this peer-to-peer interaction that is the source of innovation and growth.

Moreover, the idea of "rational individuals" is not what Adam Smith said created the invisible hand. Instead, Adam Smith thought: "It is human nature to exchange not only goods but also ideas, assistance, and favors…it is these exchanges that guide men to create solutions for the good of the community." Interestingly, Karl Marx said something similar, namely that society is the sum of all our social relationships.

The norms of society, the solutions for society, come from peer-to-peer communication – not just from markets and not just from independent individual reasoning. Financial systems and investors should focus much more on interactions between individuals, and not just on the behavior of individuals. Until recently, we didn't have data to examine these interactions, nor did we have the right sort of mathematics model networks of peer-to-peer interaction. Now we have both the math and the data to better understand and govern ourselves and can begin to understand ourselves more clearly, and invest more reliably.

References

[1] Mann, A. (2016) Core Concept: Computational social science, PNAS January 19, 2016 113 (3) 468-470; https://doi.org/10.1073/pnas.1524881113, http://www.pnas.org/content/113/3/468.full

[2] Pentland, A. (2012) The New Science of Building Great Teams, Harvard Business Review, April 2012 https://hbr.org/2012/04/the-new-science-of-building-great-teams

[3] Buchanan, M. (2009) Behavioral Science: Secret Signals, *Nature* **457**, 528-530 (2009) | doi:10.1038/457528a http://www.nature.com/news/2009/090128/full/457528a.html

[4] Bogomolov, A., Lepri, B., Staiano, J., Oliver, N., Pianesi, F., and Pentland, A. (2014). Once upon a crime: Towards crime prediction from demographics and mobile phone data. Proceedings of the 16th International Conference on Multimodal Interaction (ICMI 2014), pp. 427-434.

[5] Jean, N., Burke, M., Xie, M., Davis, W.M., Lobell, D.B., and Ermon, S. (2016). Combining satellite imagery and machine learning to predict poverty. *Science*. 353(6301), pp. 790-794.

[6] Independent Expert Advisory Group on a Data Revolution for Sustainable Development. (2014). *A world that counts: Mobilising the data devolution for sustainable development*.

[7] Lu, X., Bengtsson, L., and Holme, P. (2012). Predictability of population displacement after the 2010 Haiti earthquake. *PNAS*. 109:11576-81.

[8] Pastor-Escuerdo, D., Torres Fernandez, Y., Bauer, J.M., Wadhwa, A., Castro-Correa, C., Romanoff, L., Lee, J.G., Rutherford, A., Frias-Martinez, V., Oliver, N., Frias-Martinez, E., and Luengo-Oroz, M. (2014). Flooding through the lens of mobile phone activity. In IEEE Global Humanitarian Technology Conference (GHTC 2014).

[9] Ruktanonchai, N.W., Bhavnani, D., Sorichetta, A., Bengtsoon, L., Carter, K.H., Cordoba, R.C., Le Menach, A., Lu, X., Wetter, E., zu Erbach-Schoenberg, E., and Tatem, A.J. (2016). Census-derived migration data as a tool for informing malaria elimination policy. *Malaria Journal*. 15:273.

[10] Smith-Clarke, C., Mashhadi, A., and Capra, L. Poverty on the cheap: Estimating poverty maps using aggregated mobile communication. (2014) In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI 2014), pp. 511-520.

[11] Soto, V., Frias-Martinez, V., Virseda, J., and Frias-Martinez, E. (2011). Prediction of socioeconomic levels using cell phone records. In Proceedings of the International Conference on User Modeling, Adaptation, and Personalization (UMAP 2011), pp. 377-388.

[12] Steele, J.E., Sundsøy, P.R., Pezzulo, C., Alegana, V.A., Bird, T.J., Blumenstock, J., Bjelland, J., Engø-Monsen, K., de Montjoye, Y.-A., Iqbal, A.M., Hadiuzzaman, K.N., Lu, X., Wetter, E., Tatem, A.J., and Bengtsson, L. (2017). Mapping poverty using mobile phone and satellite data. *Journal of the Royal Society Interface* 14(127).

[13] Traunmueller, M., Quattrone, G., and Capra, L. (2014). Mining mobile phone data to investigate urban crime theories at scale. In Proceedings of the International Conference on Social Informatics (SocInfo 2014), pp. 396-411.

[14] Wang, H., Kifer, D., Graif, C., and Li, Z. (2016). Crime rate inference with big data. In Proceedings of the ACM SIGKDD Conference on Knowledge Discovery and Data Mining (KDD 2016), pp. 635-644.

[15] Wesolowski, A., Eagle, N., Tatem A.J., Smith, D.L., Noor, A.M., Snow, R.W., Buckee, C.O. (2012). Quantifying the impact of human mobility on malaria. *Science*. 338(6104), pp. 267-270.

[16] Wesolowski, A., Metcalf, C.J.E., Eagle, N., Kombich, J., Grenfell, B.T., Bjørnstad, O.N., Lessler, J., Tatem, A.J., and Buckee, C.O. (2015). Quantifying seasonal population fluxes driving rubella transmission dynamics using mobile phone data. *PNAS*. 112(35), pp. 11114-11119.

[17] Buera, F., Oberfield, E., (2016) The Global Diffusion of Ideas, NBER Working Paper No. 21844

[18] Alvarez, F., Buera, F., Lucas, R. (2013) Idea Flows, Economic Growth, and Trade, NBER Working Paper No. 19667

[19] Improving official statistics in emerging markets using machine learning and mobile phone data, (2017) Eaman Jahani, Pål Sundsøy, Johannes Bjelland, Linus Bengtsson, Alex 'Sandy' Pentland and Yves-Alexandre de Montjoye, *EPJ Data Science,* 2017 **6**:3, https://doi.org/10.1140/epjds/s13688-017-0099-3

[20] The predictability of consumer visitation patterns, (2013) Coco Krumme,, Alejandro Llorente, , Manuel Cebrian, Alex ("Sandy") Pentland Esteban Moro, *Scientific Reports* volume3, Article number: 1645

[21] Sensing the" health state" of a community, (2012), Anmol Madan, Manuel Cebrian, Sai Moturu, Katayoun Farrahi, IEEE Pervasive Computing, Vol. 11, No. 4, pp. 36-45

[22] Money Walks: Implicit Mobility Behavior and Financial Well-Being (2015) Vivek Singh, Burcin Bozkaya, Alex Pentland, PloS, https://doi.org/10.1371/journal.pone.0136628

[23] Urban characteristics attributable to density-driven tie formation, Wei Pan, Gourab Ghoshal,, Coco Krumme, Manuel Cebrian, and Alex Pentland, *Nature Communications, V*olume4, Article number: 1961 (2013)

[24] Social Physics: How social networks can make us smarter (2015) Alex Pentland, Penguin Press.

[25} Social Bridges in Urban Purchase Behavior, (2018) Xiaowen Dong, Yoshi Suhara, Burcin Bozkaya, Vivek Singh, Bruno Lepri, Alex Pentland, ACM Transactions on Intelligent Systems and Technology (TIST), Vol 9, No. 3, Article 33.

[26] A Computational Approach to Urban Economics, Shi Chong, Mohsen Baharmi, Alex Pentland, *in preparation,* 2018

[27] Research on the Relationship Between Entrepreneurship Diversity and Performance of Regional Innovation Networks——Micro data analysis based on 3255 Chinese incubators. (2017), Yuan,X.F., Qiu, Y.J., Wei, Y. , & Liu, S.Y., Theory&Practice, 2017(9),152-155.

[28] Collective Learning in China's Regional Economic Development (2018), Jian Gao, Bogang Jun, Alex Pentland, Tao Zhou, and Cesar A. Hidalgo, https://arxiv.org/abs/1703.01369. (2016). Human Collective Intelligence as Distributed Bayesian Inference.

[29] Peter Krafft, Julia Zheng, Wei Pan, Nicolás Della Penna, Yaniv Altshuler, Erez Shmueli, Joshua B. Tenenbaum, and Alex Pentland *(2016), arXiv:1608.01987*.

[30] http://trust.mit.edu

[31] Mapping the Risk-Utility Landscale: Mobile Data for Sustainable Development & Humanitarian Action, see http://unglobalpulse.org