3-2020

Quantum (and More General) Models of Research Collaboration

Oscar Galindo
The University of Texas at El Paso, ogalindomo@miners.utep.edu

Miroslav Svitek
Czech Technical University in Prague, svitek@fd.cvut.cz

Vladik Kreinovich
The University of Texas at El Paso, vladik@utep.edu

Comments:
Technical Report: UTEP-CS-20-18
Published in *Asian Journal of Economics and Banking*, 2020, Vol. 4, No. 1, pp. 77-86.

Quantum (and More General) Models of Research Collaboration

Oscar Galindo[1], Miroslav Svitek[2], and Vladik Kreinovich[1]
[1]Department of Computer Science
Universtity of Texas at El Paso
500 W. University
El Paso, TX 79968, USA
ogalindomo@miners.utep.edu, vladik@utep.edu
[2]Faculty of Transportation Sciences
Czech Technical University in Prague
Konviktska 20
CZ-110 00 Prague 1, Czech Republic
svitek@fd.cvut.cz

Abstract

In the last decades, several papers have shown that quantum techniques can be successful in describing not only events in the micro-scale physical world – for which they were originally invented – but also in describing social phenomena, e.g., different economic processes. In our previous paper, we provide an explanation for this somewhat surprising successes. In this paper, we extend this explanation and show that quantum (and more general) techniques can also be used to model research collaboration.

1 Formulation of the Problem

What is the problem. It is well known that when researchers collaborate, their productivity usually increases: together, they generate more results and more applications than when they work on their own. Sometimes, this increase is significant, sometimes, it is small, and sometimes, the two people simply do not match, and their attempts to collaborate turn out to be counter-productive. To enhance productivity, it is therefore desirable to predict how well a group of researchers can work together. For this prediction, we can use the experience of their past collaboration with each others and with other researchers. To make such a prediction, we need a good model of such collaboration efficiency.

The existing model do not always provide a good prediction, so new models are needed.

How can we approach this problem. We need a quantitative model to describe a social phenomenon. In general, quantitative methods appeared in social sciences much later than in natural science – such as physics. As a result, from the mathematical viewpoint, we can claim that quantitative models in social sciences are "behind" physics models. For example, stochastic differential equations have been, in effect, used in physics for many decades – e.g., to describe the Brownian motion – while in social sciences, such models appeared only a few decades ago, to predict the cost of financial derivatives.

From this viewpoint, to make social models more adequate, a natural idea is to try to use mathematics behind more recent physics models. One formalism that is been actively used in physics is the formalism of quantum mechanics. It is thus reasonable to try to use quantum techniques to describe social phenomena. This idea has been indeed successfully tried; see, e.g., [1, 3, 4, 10, 11, 12] and references therein.

At first glance, these successes may sound accidental – after all, quantum phenomena in physics are very different from social phenomena. However, as we have shown in [12], there is a solid explanation behind these successes – namely, a detailed stochastic analysis of the corresponding social phenomena shows that quantum-type formulas indeed appear as a reasonable first approximation to these phenomena. That paper also provides formulas for – possible more accurate – next approximations.

What we do in this paper. In this paper, we extend the analysis from [12] to the collaboration phenomenon. Namely, we show that this analysis leads, in the first approximation, to a quantum-type model of collaboration phenomena. We also provide ideas for next (post-quantum) approximations.

Comment. In this paper, we focus on research collaboration, since the results of this type of collaboration can be naturally quantified. However, we believe that similar techniques can be useful in describing other types of collaboration as well – e.g., between musicians in an orchestra or between athletes in a sports team, provided that we (find and) use an appropriate numerical measure for the successfulness of such collaboration.

2 Analysis of the Problem

Stochastic character of research collaboration. Research activity is a difficult-to-predict phenomenon, probably the hardest-to-predict: if it was easy to predict the results, there would be no need for creative collaboration. The results of a creative process depends on many difficult-to-predict factors: an idea comes to mind, an article happens to appear in a journal that deals with a similar problem, a new analogy suddenly pops into mind. All these factors are largely random.

Thus, from the mathematical viewpoint, a person's creative activity can be described as a random process depending on many independent factors.

None of these factors is dominant: in some cases, a new paper prompted the insight; in other cases, it was some piece of art – for example, Einstein famously said that Dostoyevsky inspired him even more than Gauss. We can therefore conclude that the individual's research productivity is the result of a joint effect of many relatively small independent factors.

It is known that, under reasonable assumptions, such a joint effect is reasonably well described by a Gaussian (normal) distribution. To be more precise, the corresponding *Central Limit Theorem* (see, e.g., [9]) states that when the number N of such small factors tends to infinity, the probability distribution of their summary effect tends to Gaussian – which means exactly that if N is large, the corresponding probability distribution is close to Gaussian.

This description applies both to the state x_i of each researcher i, and to the vector $x = (x_1, \ldots, x_n)$ describing the state of each of the researchers from a given group. So, we can conclude that the state of all the researchers can be described by a multi-D Gaussian process. If instead of the original state, we consider the *deviations* from the original state – the only thing that matters in research – we can conclude that the mean value of this newly considered state is 0.

Vector descriptions. Some factors affecting productivity are individual to each researcher, others may be common to several researchers. As a result, the corresponding random variables are correlated.

It is known that a generic n-dimensional multi-D Gaussian distribution with 0 mean can be described as a linear combination of n independent standard Gaussian random variables, with 0 mean and standard deviation 1. In other words, each random variable x_i can be described as $x_i = a_{i1} \cdot e^{(1)} + \ldots + a_{in} \cdot e^{(n)}$ for some real numbers $a_{i1}, \ldots, a_{in}$, where $e^{(i)}$ are independent Gaussian variables for which the mean $E\left[e^{(i)}\right]$ is equal to 0 and the standard deviation $\sigma\left[e^{(i)}\right]$ is equal to 1.

Thus, the research activity of each individual i can be described by an n-dimensional vector $a_i = (a_{i1}, \ldots, a_{in})$. The overall productivity of a person can be estimated as a mean-square value of the corresponding deviation, i.e., since $E[x_i] = 0$, as $V_i = E[x_i^2]$. From the above formula for x_i, we can conclude that

$$V_i = \|a_i\|^2 \stackrel{\text{def}}{=} \sum_{j=1}^{n} a_{ij}^2.$$

Need for an approximate description. For large n, we need to represent each researcher by a corresponding n-dimensional vector. Science is a collective enterprise, involving thousands of people. So, to represent each researcher exactly, we would need to describe thousands of parameters. All these parameters need to be determined experimentally, but usually, we do not have enough information about each individual researcher to find all these parameters.

So, realistically, we have to ignore some of the original parameters. Thus, a natural idea is to select some number $k \ll n$ and describe each researchers by a smaller-dimensional (namely, k-dimensional) vector $A_i = (A_{i1}, \ldots, A_{ik})$.

Quantum description as a particular case of a vector description. In particular, for $k = 2$, each researcher i is characterized by a 2-D Vector $A_i = (A_{i1}, A_{i2})$, and the researcher's productivity is characterized by the value

$$V_i = A_{i1}^2 + A_{i2}^2.$$

One of the possible algebraic interpretations of a 2-D space is as a space of all complex numbers. Thus, it is natural to characterize each researcher by a complex number $C_i = A_{i1} + \mathrm{i} \cdot A_{i2}$, where $\mathrm{i} \stackrel{\mathrm{def}}{=} \sqrt{-1}$. In this case, $V_i = |C_i|^2$.

From the mathematical viewpoint, this is exactly what quantum descriptions are about – using complex numbers to describe the states, with the square of the absolute values of the corresponding quantum numbers to describe the observable quantity (in quantum case, probability of an event).

Need to describe collaboration. So far, we have described the state (and the productivity) of an individual researcher. When some of the researchers work together, their state and their productivity changes. In general, if we have m researchers, we can have all possible groups working together, i.e., all possible non-empty subsets $S \subseteq \{1, \ldots, m\}$.

Suppose that we know the states $A_1, \ldots, A_r$ of all the researchers. Based on these states, we can estimate the productivity $\|A_i\|^2$ of each researcher i.

We would like to predict the state A_S of all possible groups S – and thus, predict the productivity $\|A_S\|^2$ of each such group.

How can we describe collaboration. Let us start with an informal description. Each group S can be characterized by its characteristic function χ_S, i.e., by assigning, to each researcher i, a value $\chi_S(i)$ which is equal to 1 if the researcher i is a member of this group and 0 if not. In there terms, the desired state A_S is a function of n binary (0-or-1) variables $\chi_S(1), \ldots, \chi_S(n)$: $A_S = F(\chi_S(1), \ldots, \chi_S(m))$.

In many application areas, the dependencies $f(x_1, \ldots, x_m)$ are smooth, so we can expand them in Taylor series

$$f(x_1, \ldots, x_m) = a_0 + \sum_{i=1}^{m} a_i \cdot x_i + \sum_{i_1, i_2 = 1}^{m} a_{i_1 i_2} \cdot x_{i_1} \cdot x_{i_2} + \ldots, \tag{1}$$

and, if needed, approximate this dependence by keeping only the few first terms in this expansion. In many cases, already the linear approximation

$$f(x_1, \ldots, x_m) = a_0 + \sum_{i=1}^{m} a_i \cdot x_i \tag{2}$$

works reasonably well; see, e.g., [2].

In our case, the variables are discrete, so we cannot talk about smooth dependence. However, it turns out that a similar formula (1) is applicable in this case as well: namely, every function of m binary variables can be described by a similar formula; see, e.g., [5, 6, 7, 8]. To be more precise, for binary

variables, the square of each variable coincides with its original value $x_i^2 = x_i$, so it is sufficient to only consider coefficients $a_{i_1\ldots i_d}$ in which the indices do not repeat. Thus, we get a form

$$f(x_1,\ldots,x_m) = a_0 + \sum_{i=1}^{m} a_i \cdot x_i + \sum_{i<i'} a_{i_1 i_2} \cdot x_{i_1} \cdot x_{i_2} + \sum_{i_1<i_2<i_3} a_{i_1 i_2 i_3} \cdot x_{i_1} \cdot x_{i_2} \cdot x_{i_3} + \ldots \quad (3)$$

The transformation that maps each function of n binary variables into the corresponding coefficients a_0, a_i, etc., is known as the *Möbius transform for partially ordered sets*; see, e.g., [8].

This dependence holds for each component A_{Sj} of the desired state vector $A_S = (A_{S1},\ldots,A_{Sn})$. In this case, $x_i = \chi_S(i)$, and the corresponding product $\chi_S(i_1)\cdot\ldots\cdot\chi_S(i_d)$ is different from 0 if and only if all the researchers $i_1,\ldots,i_d$ belong to the group S, i.e., if and only if $\{i_1,\ldots,i_d\} \subseteq S$. In this case, the product $\chi_S(i_1)\cdot\ldots\cdot\chi_S(i_d)$ is equal to 1. So, for the component A_{Sj}, the formula (3) takes the following form:

$$A_{Sj} = a_{0j} + \sum_{i\in S} a_{ij} + \sum_{i_1,i_2:\{i_1,i_2\}\subseteq S} a_{i_1 i_2 j} + \sum_{i_1,i_2,i_3:\{i_1,i_2,i_3\}\subseteq S} a_{i_1 i_2 i_3 j} + \ldots \quad (4)$$

The case $S = \emptyset$ means that no one is doing anything. In this case, the productivity is 0, so we should have $A_{Sj} = 0$. For $S = \emptyset$, the formula (4) turns into $A_{Sj} = a_{0j}$, so we conclude that $a_{0j} = 0$. Thus, the formula (4) takes a simplified form

$$A_{Sj} = \sum_{i\in S} a_{ij} + \sum_{i_1,i_2:\{i_1,i_2\}\subseteq S} a_{i_1 i_2 j} + \sum_{i_1,i_2,i_3:\{i_1,i_2,i_3\}\subseteq S} a_{i_1 i_2 i_3 j} + \ldots \quad (5)$$

What if we have a researcher working on his/her own? In this case, $S = \{i\}$, and the right-hand side of the formula (5) turns into a_{ij}. On the other hand, in this case, A_{Sj} is the j-th component A_{ij} of the vector A_i describing the state of the i-th researcher. Thus, we conclude that $a_{ij} = A_{ij}$ and so, the formula (5) takes the form

$$A_{Sj} = \sum_{i\in S} A_{ij} + \sum_{i_1,i_2:\{i_1,i_2\}\subseteq S} a_{i_1 i_2 j} + \sum_{i_1,i_2,i_3:\{i_1,i_2,i_3\}\subseteq S} a_{i_1 i_2 i_3 j} + \ldots \quad (6)$$

We can describe this in vector terms, if we introduce vectors

$$A_{i_1\ldots i_d} \stackrel{\text{def}}{=} (a_{i_1\ldots i_d 1},\ldots,a_{i_1\ldots i_d k});$$

then, the formula (6) takes the form:

$$A_S = \sum_{i\in S} A_i + \sum_{i_1,i_2:\{i_1,i_2\}\subseteq S} A_{i_1 i_2} + \sum_{i_1,i_2,i_3:\{i_1,i_2,i_3\}\subseteq S} A_{i_1 i_2 i_3} + \ldots \quad (7)$$

In particular, in the first approximation, when we only keep linear terms, we get

$$A_{Sj} = \sum_{i \in S} A_{ij}, \tag{8}$$

i.e., in terms of vectors,

$$A_S = \sum_{i \in S} A_i. \tag{9}$$

In the 2-D case, if we describe 2-D vectors A_i, $A_{i_1 i_2}$, ..., and A_S by complex numbers C_i, $C_{i_1 i_2}$, ..., and C_S, we get

$$C_S = \sum_{i \in S} C_i + \sum_{i_1, i_2 : \{i_1, i_2\} \subseteq S} C_{i_1 i_2} + \sum_{i_1, i_2, i_3 : \{i_1, i_2, i_3\} \subseteq S} C_{i_1 i_2 i_3} + \ldots \tag{10}$$

In the first approximation, when we only keep linear terms, we get:

$$C_S = \sum_{i \in S} C_i. \tag{11}$$

Thus, we arrive at the following model for describing research collaboration.

3 Resulting Model and Its Analysis

Resulting model: general case. To describe research collaboration, we need to select two parameters:

- the dimension k of the vectors describing the state of each research (and each group), and
- the order d of the terms that we use to describe collaboration.

The larger k and the larger d, the more accurate the description.

In this model, the state of each researcher is described by a k-dimensional vector $A_i = (A_{i1}, \ldots, A_{ik})$. In these terms, the productivity V_i of an individual researcher is equal to $V_i = \|A_i\|^2$, where, for each vector $v = (v_1, \ldots, v_k)$, the value $\|v\| \stackrel{\text{def}}{=} \sqrt{v_1^2 + \ldots + v_k^2}$ denotes its length.

In particular, when $k = 2$, we can describe this state $A_i = (A_{i1}, A_{i2})$ as a complex number $C_i = A_{i1} + \mathrm{i} \cdot A_{i2}$. In terms of this complex number, the length of the vector A_i is then equal to $|C_i|^2$.

For each group S of collaborating researchers, the vector A_S that describes the state of this group has the form:

$$A_S = \sum_{i \in S} A_i + \sum_{i_1, i_2 : \{i_1, i_2\} \subseteq S} A_{i_1 i_2} + \ldots + \sum_{i_1, \ldots, i_d : \{i_1, \ldots, i_d\} \subseteq S} A_{i_1 \ldots i_d}, \tag{12}$$

for appropriate auxiliary vectors $A_{i_1 i_2}$, ..., $A_{i_1 \ldots i_d}$. The productivity V_S of the group S is equal to $\|A_S\|^2$.

In particular, in the complex-valued case $k = 2$, we get

$$C_S = \sum_{i \in S} C_i + \sum_{i_1, i_2 : \{i_1, i_2\} \subseteq S} C_{i_1 i_2} + \ldots + \sum_{i_1, \ldots, i_d : \{i_1, \ldots, i_d\} \subseteq S} C_{i_1 \ldots i_d}, \qquad (13)$$

and the productivity is equal to $|C_S|^2$.

It is necessary to select small k and d. The larger the values of k and n, the more accurate the resulting model – but, on the other hand, the more parameters we will need to describe this model, and we usually do not have enough observations to determine too many parameters. Thus, we need to concentrate on models with small k and d, for which the number of parameters is still reasonable.

Let us start with the smallest possible values, and, if the resulting oversimplified model is not too realistic, let us see how to increase k or d to make the model more realistic while keeping the number of parameters reasonable.

What is the simplest case: description and analysis. The simplest case is when we select the smallest possible values $k = 1$ and $d = 1$. In this case, the state of each researcher i is described by a single number A_i, with productivity $V_i = |A_i|^2$, the state A_S of a group S is equal to the sum $A_S = \sum_{i \in S} A_i$, and the productivity of the group is equal to $|V_i|^2$.

In particular, for a group $S = \{1, 2\}$ consisting of two researchers, we have $V_S = |A_1 + A_2|^2$. Here, in terms of the productivity V_i, the state A_i of the i-th researcher has the form $A_i = \pm\sqrt{V_i}$, so $V_S = |\pm\sqrt{V_1} \pm \sqrt{V_2}|^2$. Depending on the signs of A_i, we have two possible options:

- if both states A_1 and A_2 have the same sign, then

$$V_S = (\sqrt{V_1} + \sqrt{V_2})^2 = V_1 + V_2 + 2\sqrt{V_1} \cdot \sqrt{V_2};$$

- if the states A_1 and A_2 have different signs, then

$$V_S = (\sqrt{V_1} - \sqrt{V_2})^2 = V_1 + V_2 - 2\sqrt{V_1} \cdot \sqrt{V_2}.$$

Thus, in this simplified model, we capture two cases:

- when the researchers most successfully collaborate, and
- when the researchers are so incompatible with each other that their overall productivity as a group is smaller than the productivity of one of them.

Limitations of the simplest model and need for a more complex one. In reality, in addition to these two extreme cases, we have the whole spectrum of possible collaboration success. To capture this spectrum, we need to consider a more adequate model, i.e., we need to either increase k or increase d (or increase both).

Which is the next simplest model? The next simplest model is when we:

- either increase k by one, leaving d unchanged, or
- or increase d by one, leaving k unchanged.

In the first case, when we take $k = 2$ and $d = 1$, to describe the situation of m possibly collaborating researchers, we need to know the values of m 2-dimensional vectors. So, in this case, overall, we need $2m$ numerical parameters.

In the second case, when we take $k = 1$ and $d = 2$, we need to describe m numbers C_i corresponding to linear order terms, and $\frac{m \cdot (m-1)}{2}$ values $C_{i_1 i_2}$ corresponding to second order terms in the formula (10) – exactly as many as there are pairs (i_1, i_2) with $i_1 < i_2$. So, in this case, overall, we need $m + \frac{m \cdot (m-1)}{2} = \frac{m \cdot (m+1)}{2}$ parameters.

The second number of parameters is larger if $\frac{m \cdot (m+1)}{2} > 2m$, i.e., when $\frac{m+1}{2} > 2$, i.e., when $m + 1 > 4$ and $m > 3$. So, for each group consisting of at least 4 researchers, the first model (with $k = 2$ and $d = 1$) requires fewer parameters and is, thus, much simpler. Because of this, in the following text, we will consider the model with $k = 2$ and $d = 1$.

Let us consider this next simplest model. In this next simplest model, each participant is described by a 2-D vector – i.e., equivalently, by a complex number C_i. The productivity of each group S is equal to $P_S = |C_S|^2$, where

$$C_S = \sum_{i \in S} C_i. \tag{14}$$

Thus, we have

$$P_S = |C_S|^2 = \sum_{i \in S} |C_i|^2 + 2 \sum_{i_1 < i_2} |C_i| \cdot |C_j| \cdot \cos(\alpha_{ij}), \tag{15}$$

where α_{ij} is the angle between the 2-D vectors C_i and C_j. In particular, for each participant i, his/her productivity P_i is equal to $|C_i|^2$. So, in terms of individual productivity values, $|C_i| = \sqrt{V_i}$ and the formula (15) takes the form:

$$P_S = \sum_{i \in S} P_i + 2 \sum_{i_1 < i_2} \sqrt{P_i} \cdot \sqrt{P_j} \cdot \cos(\alpha_{ij}). \tag{16}$$

Our preliminary analysis shows that this model describes the collaboration between folks reasonably well – at least on the qualitative level. When the cosine is positive, the group's productivity is much larger than the combined productivity of all its members. For example, if all the angles are 0s, and all the cosines are equal to 1, then for a group of m members its productivity grows as m^2, while the sum of productivity values grows only as m.

On the other hand, if the cosines are negative, the overall productivity is smaller than it would be if everyone worked on their own. In this case, researchers are clearly not compatible, and collaboration does not make sense.

Possible future work. It would be nice to compare, on a quantitative level, how well collaboration results can be described by this model. It this model turns out to be adequate, we can then use it to decide how to group people into collaborating teams – so as the overall productivity of all these teams is the largest possible.

Of course, even if the model works well, it will be only approximate – since practically all models of real-life situations, especially models involving human behavior, are approximate. For applications requiring higher accuracy, it would then be natural to look for a more accurate model – e.g., to consider larger values of parameters k and n.

4 Acknowledgments

This work was supported in part by the Project AI & Reasoning CZ.02.1.01/0.0/0.0/15_003/0000466, by the European Regional Development Fund, and by the US National Science Foundation grants 1623190 (A Model of Change for Preparing a New Generation for Professional Practice in Computer Science) and HRD-1242122 (Cyber-ShARE Center of Excellence).

This work was partially done when Miroslav Svitek was a visiting researcher at the University of Texas at El Paso.

References

[1] B. E. Baaquie, *Quantum Finance: Path Integrals and Hamiltonians for Options and Interest Rates*, Camridge University Press, New York, 2004.

[2] R. Feynman, R. Leighton, and M. Sands, *The Feynman Lectures on Physics*, Addison Wesley, Boston, Massachusetts, 2005.

[3] E. Haven abnd A. Khrennikov, *Quantum Social Science*, Cambridge University Press, Cambridge, UK, 2013.

[4] V. Kreinovich, H. T. Nguyen, and S. Sriboonchitta, "Quantum Ideas in Economics Beyond Quantum Econometrics", In: Ly H. Anh, Le Si Dong, V. kreinovich, and Nguyen Ngoc Thach (eds.), *Econometrics for Financial Applications*, Springer Verlag, Cham, Switzerland, 2018, pp. 146–151.

[5] J. Nava and V. Kreinovich, "Equivalence of Gian-Carlo Rota Poset Approach and Taylor Series Approach Extended to Variant Ligands", *Journal of Uncertain Systems*, 2011, Vol. 5, No. 2, pp. 111–118.

[6] J. Nava and V. Kreinovich, *Algorithmic Aspects of Analysis, Prediction, and Control in Science and Engineering: An Approach Based on Symmetry and Similarity*, Springer Verlag, Berlin, Heidelberg, 2015.

[7] J. Nava, V. Kreinovich, G. Restrepo, and D. J. Klein, "Discrete Taylor Series as a Simple Way to Predict Properties of Chemical Substances like

Benzenes and Cubanes", *Journal of Uncertain Systems*, 2010, Vol. 4, No. 4, pp. 270–290.

[8] G.-C. Rota, "On the foundations of combinatorial theory I. Theory of Möbius functions", *Zeit. Wahrscheinlichkeitstheorie*, 1964, Vol. 2, pp. 340–368.

[9] D. J. Sheskin, *Handbook of Parametric and Nonparametric Statistical Procedures*, Chapman and Hall/CRC, Boca Raton, Florida, 2011.

[10] M. Svítek, *Quantum System Theory: Principles and Applications*, VDM Verlag, Saarbrucken, Germany, 2010.

[11] M. Svítek, "Towards complex system theory", *Neural Networl World*, 2015, Vol. 15, No. 1, pp. 5–33.

[12] M. Svitek, O. Kosheleva, V. Kreinovich, and Thach Ngoc Nguyen, "Why quantum (wave probability) models are a good description of many non-quantum complex systems, and how to go beyond quantum models", In: V. Kreinovich, Nguyen Ngoc Thach, Nguyen Duc Trung, and Dang Van Thanh (eds.), *Beyond Traditional Probabilistic Methods in Economics*, Springer, Cham, Switzerland, 2019, pp. 168–175.