Progress in Physics (33)

Outreach: Can Physics Cross Boundaries?

Jean-Pierre Eckmann, Département de Physique Théorique et Section de Mathématiques, Université de Genève

Recently, physical thinking has been making progress in domains at which it did not aim originally. In this contribution, I want to sketch some examples of what can be done. The method consists of finding systems which originate in complex or complicated structure or dynamics, and which can profit from questions physicists ask. The examples I want to present comprise biology, language and the World-wide-web (WWW).

I find it fascinating that relatively simple methods, questions, and techniques from the exact sciences seem to be able to shed new light, and also new insight, into structures which are mostly self-generated. I want to suggest and illustrate that questions outside physics proper can be developed fruitfully by physicists. My story is neither totally new (see, e.g., [14]) nor as revolutionary as it may seem. I just want to convey my interest and pleasure in addressing "esoteric" questions with the tools of mathematics and physics.

The discussion will be in the subject of "network theory," and I first summarize some of its literature [1, 12] : With the advent of powerful computers on every scientist's desk, it has become easy to analyze large data sets. These data sets come often, and quite naturally, in the form of large networks (graphs, directed or undirected), where the nodes of the graph are certain objects, and the edges are certain binary relations between them. For example, the nodes could be individual researchers, and the links could signify that they either co-author a paper, or cite each other. Other examples are pages and links in the WWW, which connect two pages; airports and connections provided by commercial airlines; words and links between these words and their definition in a dictionary. I will call such graphs real-life graphs [1]. Experimental automation, and the availability of large databases through the internet provide many interesting networks for analysis. The most useful ones are obtained in collaboration with experimental scientists.

Continuing a long tradition in statistical physics, the studies of large networks often concentrate on their statistical properties. Erdős and Rényi described a set of random graphs which are built as follows [3]: Take N nodes (N very large) and assume that the mean degree (number of links coming out of a node) is $k > 0$, independently of N. Then, paraphrasing Erdős and Rényi, one can make two statements:

1. Such a graph looks locally like a tree (i.e., it has very few loops, and these loops are all very long) [4].
2. The expected number of triangles is $k^3/6$, (i.e., this number does *not* grow with N). (Longer loops are also rare [2].)

1 One may legitimately ask why only binary relations seem important, but I will argue later that triangles in these graphs play the role of three-body-interactions and are the main indicators of semantic contexts.

2 It is actually quite easy to prove these statements, although, at first, they certainly seem totally anti-intuitive.

The first surprise was the discovery that real-life graphs are *not* random in the above sense. In contrast to general results on random graphs, the graphs of "affinities" or "connections" between "authors," "entities" have very specific general properties, namely power law behavior over several decades [1]. By this, one means the statistics of the number $N(j)$ of nodes which are attached to exactly j others (the degree of the node). In formulas, $N(j) \approx const. j^{-\gamma}$ for large j. The point here is that the decay is a power law, and not an exponential, pointing to the important feature of nodes in the graph with very many connections, many more than a Gaussian, or Poisson distribution would allow for. In many real-life graphs γ takes a value between 2 and 3.

What this means is that there are a few nodes which have a really high degree. For example, in studying the connections in Twitter, one finds that there are a few nodes with over 100'000 "friends" (these are usually politicians, perhaps also singers), the interesting question here is whether they are friends (the singers) or whether they think they have friends (the politicians...).

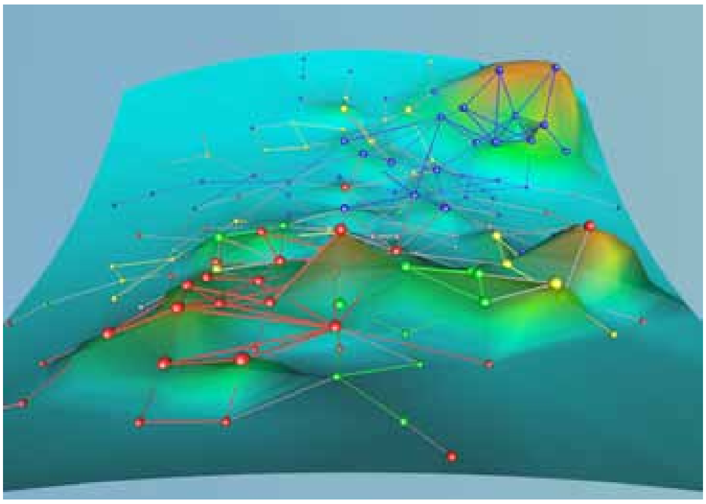
Many studies then concentrate on the dynamics of how such networks come into being. This is usually called the "preferential attachment" [3] problem [2], namely the idea that the networks build up in time, and that people have a tendency to connect to well-known other people (or services). These models have successfully explained how long-range (scale-free) properties of graphs come about.

Another important aspect of network studies is summarized under the term of "clustering coefficient" [18]. In contrast to the power laws described above, this is a local property of any graph. In mathematical terms, if a node n has j neighbors, then you count the number t of triangles which have the node n as a corner. Obviously, there cannot be more than $T(j) = j(j-1)/2$ such triangles, and the clustering coefficient is defined as $t/T(j)$, which is a number between 0 and 1. A high clustering coefficient means that many of the possible triangles are actually realized.

When I started to study real-life graphs [7], I was puzzled by the abundance of triangles, which appear orders of magnitude above $\mathcal{O}(k^3)$ predicted by Erdős and Rényi. What does this mean? It soon turned out that triangles play a strong semantic role. In other words, in all studies of this type, one can attach meaning to this abundance.

The first case where we discovered this phenomenon was the set of links in the WWW [7]. While there are many links which seem irrelevant, we found that those links which form triangles relate to common interests of the owners of the pages involved. Carrying this idea further, we found that triangles of connections among neurons of *C. Elegans* (a little worm with 302 neurons) organize their function. Another example is provided by triangles of e-mail messages sent between people, and this determines their social grouping [8].

3 also called "rich get richer"

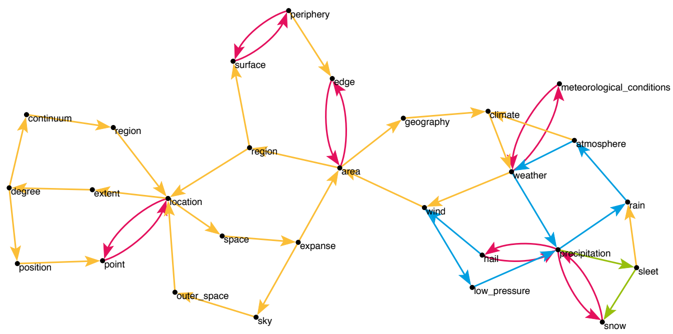


The neurons of C. Elegans and their connections: Height is clustering coefficient, color codes function (motor, sensory,...)

My final example is the appearance of loops in the definitions of words in a dictionary [10]. Here, the graph is formed by arrows pointing from each word (actually nouns) to the words in their definition. One finds not only triangles, but also bi-angles, back-and-forth, (which are synonyms) and also longer loops. When the loops are too long, there might be a jump in interpretation, such as at the broken arrow in

> *railcar* ⇒ *rails* ⇒ *bar* ⇒ *weapon* ⇒ *instrument* ⇒
> *skill* ⇏ *train* ⇒ *railcar*.

But the surprising fact is that the medium size loops are semantically coherent, and their words form the core of the language. (We checked that by comparing the core-words to those of [13].) And furthermore, as a bonus, one can show that they are related to the historical appearance of new concepts in language [9].



The structuring of words in a dictionary.

What fascinates me in all this is that questions and methods from physics can shed new light on several systems with complex dynamics or structure. The most accessible among such systems are those where sub-units are assembled without a master building plan.

The common structure which appears in such studies is that the more "interesting dynamics" (biochemistry, feedback) [16], the "meaning" (language) [5] or "mechanisms" (emergent life) are all revealed by deviations from the natural statistical structures of random assemblies. In general, the connected objects have some "deeper" identity (meaning, semantics) that curves the pathways back to some originating node, in contrast to the "rich get richer" scenario.

So far, these methods have allowed to give some insight into realms outside of physics proper. I am convinced that this outreach has an interesting future, and will penetrate further new domains.

Where does all this lead? The studies I have presented all use only the *topology* of the network, but never the *metrics*, namely the way some real-life networks are embedded in the space in which we live. Adding this component opens new possibilities, especially in studying aspects of life. This is so because many aspects of life are about spatial connections, organisms are a linked set of individual units that have to interact well and "intelligently" in order to function.

It is too early to make strong methodological statements about how mathematics and physics can play a role in studying such problems, but certainly, a few studies on neuronal networks [6, 17], or the interaction of ants in colonies [11, 15] seem to me promising beginnings in developing a methodology for studies in living systems.

References

[1] R. Albert and A.-L. Barabási. Statistical mechanics of complex networks. *Rev. Modern Phys.* **74** (2002), 47–97.

[2] A.-L. Barabási and R. Albert. Emergence of scaling in random networks. *Science* **286** (1999), 509–512.

[3] B. Bollobás. *Random graphs*, volume 73 of *Cambridge Studies in Advanced Mathematics* (Cambridge: Cambridge University Press, 2001), second edition.

[4] A. Dembo and A. Montanari. Ising models on locally tree-like graphs. *Ann. Appl. Probab.* **20** (2010), 565–592.

[5] B. Dorow, D. Widdows, K. Ling, J.-P. Eckmann, D. Sergi, and E. Moses. Using curvature and Markov clustering in graphs for lexical acquisition and word sense discrimination. In: *MEANING-2005, 2nd Workshop organized by the MEANING Project, February 3rd-4th 2005, Trento, Italy.* (2005).

[6] J.-P. Eckmann, O. Feinermann, L. Gruendlinger, E. Moses, J. Soriano, and T. Tlusty. The physics of living neural networks. *Physics Reports* **449** (2007), 54–76.

[7] J.-P. Eckmann and E. Moses. Curvature of co-links uncovers hidden thematic layers in the World Wide Web. *Proc. Natl. Acad. Sci. USA* **99** (2002), 5825–5829 (electronic).

[8] J.-P. Eckmann, E. Moses, and D. Sergi. Entropy of dialogues creates coherent structures in e-mail traffic. *Proc. Natl. Acad. Sci. USA* **101** (2004), 14333–14337 (electronic).

[9] D. Harper. *Online Etymology Dictionary.* (2010).

[10] D. Levary, J.-P. Eckmann, E. Moses, and T. Tlusty. Loops and self-reference in the construction of dictionaries. *Phys. Rev. X* **2** (2012), 031018.

[11] D. P. Mersch, A. Crespi, and L. Keller. Tracking individuals shows spatial fidelity is a key regulator of ant social organization. *Science* **340**.

[12] M. Newman, A.-L. Barabási, and D. J. Watts, eds. The structure and dynamics of networks. *Princeton Studies in Complexity* (Princeton, NJ: Princeton University Press, 2006).

[13] C. Ogden. *Basic English: a general introduction with rules and grammar* (Kegan Paul, London, 1930).

[14] L. Onsager. Nobel lecture, The motion of ions: Principles and concepts. In: *Nobel Lectures* (December 11, 1968).

[15] N. Razin, J.-P. Eckmann, and O. Feinerman. Desert ants achieve reliable recruitment across noisy interactions. *J. Royal Soc. Interface* **10** *(2013) 20130079*.

[16] S. Shen-Orr, R. Milo, S. Mangan, and U. Alon. Network motifs in the transcriptional regulation network of Escherichia coli. *Nature Genetics* **31** (2002), 64–68.

[17] T. Tlusty and J.-P. Eckmann. Remarks on bootstrap percolation in metric networks. *J. Phys. A* **42** (2009), 205004, 11.

[18] D. Watts and S. Strogatz. Collective dynamics of 'small-world' networks. *Nature* **393** (1998), 409–410.