Journal of Physics: Complexity

PERSPECTIVE • OPEN ACCESS

A perspective on complexity and networks science

To cite this article: Guido Caldarelli 2020 *J. Phys. Complex.* **1** 021001

Journal of Physics: Complexity

RECEIVED
1 April 2020

REVISED
12 May 2020

ACCEPTED FOR PUBLICATION
5 June 2020

PUBLISHED
3 July 2020

PERSPECTIVE

A perspective on complexity and networks science

Guido Caldarelli[1,2,3]

[1] IMT School for advanced studies, Piazza San Francesco 19, 55100 Lucca, Italy
[2] ECLT, Università Ca' Foscari, Ca' Bottacin, Dorsoduro 3911, Calle Crosera - 30123 Venezia, Italy
[3] London Institute for Mathematical Sciences, 35A South ST, Mayfair, London, W1K 2XF, United Kingdom

Keywords: complexity, network science, disintermediation

Abstract

Complexity and network science are nowadays used, or at least invoked, in a variety of scientific researchareas ranging from the analysis of financial systems to social structure and even to medicine. Here I explore some of the possible reasons for this success, the relationship between them and how they might be used in the future.

Complexity, (and one can probably expect it), lacks a simple definition, but somehow everybody intuitively understands this concept. More importantly, there is a large consensus on its presence and role in the everyday world [4, 26]. If we want to stick to the original formulation by Anderson [3], we immediately see in a large class of phenomena that '*more is different*', and that complexity (a new behaviour) arises whenever the number of ingredients (the particles, the agents, whatever describes the system under consideration) become very large. This explains why everybody understands what we mean when we talk about complexity. Complexity is indeed one of the features of this modern world, and it looks very familiar, since we are embedded by the increasingly growing number of relationships that we are establishing in this hyper-connected world. As a consequence of these links, and as a consequence of the information we exchange on them, the total amount of data produced and stored in the world is constantly increasing. According to a report [16], the total amount of data in the world reached 2.8 zettabytes (ZB) in 2012–or 2.8 trillion gigabytes. The volumes of data was expected to reach 40 ZB by 2020 i.e. 4×10^{22} bytes, with emerging economies accounting for an increasingly large proportion of the world's total. On the same point, following a more recent report on the 2017 review from IBM Marketing Cloud [1], we see that about 90% of the data in the world today has been created in the last two years alone, at a rate of 2.5 quintillion (10^{18}) bytes of data a day. The relations that we create can be of different natures: *physical* as it is the flow of people and goods from one country to another one; *virtual* as the information travelling on the internet on the world wide web and the services on it; *financial* with respect to the architecture of our financial systems with all the loans and credit amongst banks and firms worldwide. While on one hand this development creates new opportunities of living, trading, being informed and able to run private business, on the other hand it also poses new challenges as the recent spreading of COVID-19 clearly shows. It is then imperative to be able to master the basic principles of complexity science, with respect to the mathematical framework that allows us to deal with this unprecedented situation.

When considering this mathematical framework it appears that (maybe for historical reasons) complexity science shares many approaches with statistical physics (another discipline that deals with a large number of elements). On the other hand, very often, the applications of this theoretical machinery deal with a series of phenomena that have never been considered in the study of physics. This apparent contradiction between theory and applications is instead a great opportunity to extend quantitative and scientific approach well outside their traditional domain. Indeed, thanks to the recent developments, we are now able to quantitatively analyse a series of disciplines that, so far, accepted only a qualitative description. Amongst the various instruments of complexity science (not all of them coming from physics, but certainly used by physicists) that have been proven to be more useful for that purpose we have

- *Non-linearity* a typical characteristic of complex phenomena where the output of the system is typically unpredictable with respect to the input. It is a very general property that has been already successfully applied to a few disciplines, for example ecology [37] and biological sciences [35].

- *Adaptation* in the sense of the dynamical changes of the system (or individual or species) to the environment is one of the typical traits of the theory of evolution by Charles Darwin [22] and one of the theoretical instruments not developed in physics, but used by physicists in a series of analysis, mostly within the third instrument of complex networks [29].
- *Data science* a recent term that has been introduced to indicate a framework of theoretical and experimental instruments that allow to make sense of the above-described deluge of unstructured data.
- *Complex networks* that with respect to the above properties is rather a mathematical formalism that allows us to treat the other features in a simple formal way.

The mathematical formalism of graphs [14] is a good way to describe complex systems. If we indicate our microscopic agents as vertices and we consider their interactions as the edges, we have a natural representation of a complex system by means of a network structure [2, 18, 25, 36]. Actually, network theory plays several roles in the area of complexity. Firstly, as just mentioned, networks are the most immediate microscopic tool to describe the onset of complexity; secondly, networks are a crucial tool to model the structure and dynamics of a variety of phenomena; thirdly and finally, they allow to visualize and to filter information in these data deluge we are witnessing in these times. Of all the above features, visualization is the one that is often underrated. In fact, to see and to understand are pretty similar concepts. Consequently in almost every Indo-European language past or present, they are indicated by the same root '*weyd-' (e.g. there is the Latin *video/videor* with the meaning 'I see/I think appropriate', the ancient Greek ὁῖδα ((v)*oida*) with the meaning 'I know because I have seen'; the Sanskrit वेद (*veda*) meaning 'knowledge', up to the two modern English words 'video' and 'wise'). Even in Chinese, 'to understand' is indicated by two ideograms 明白 i.e. (*míngbai*) with the literal meaning of 'bright white'. This is just to remark how much 'visualising' and its immediate consequence 'understanding' are related and to stress that the immediacy of visualisation power is one of the most important features of complex networks. In fact, the relationship between visualisation and understanding goes far beyond network theory. The power of visualisation is widely used in data science [13, 24], climate studies [21] and so on. Obviously, such a power is strongly reduced when considering big data or very dense networks. In such cases filtering methods to enhance the most important edges work remarkably well. They range from minimal spanning trees [33] to planar graphs [40] to the use of comparative advantage [5] to evaluate economic exports in the matrices of trade between nations.

As complexity increases, one has to take into account different kinds of networks. For as for example when considering the different kind of relationsips that companies might have with each other, we need more refined versions of networks as multigraphs or multilevel networks [8, 11, 12]. In this vast scenario, we can try to select the most important issues that will attract the interest of scientist and that will have the greatest impact on society. Certainly we can predict that the distances (in the topological sense of number to edges to travel to join two vertices) in the future world, will decrease even more from their present value (not withstanding the fact that the effects of the 2020 pandemic have still to be understood). An immediate consequence of this phenomenon is the parallel growth of *disintermediation*, that is the removal of intermediate agents in the chain of distribution, discussion, and/or production. Such disintermediation, that is the possibility to establish contacts and contracts with everybody else in the world without any intermediate passage, is probably a phenomenon that deserves to be studied in great detail. A first effect (thanks to the pervasive presence of social networks on which everybody is spending more and more time) will be the inevitable consequence of getting more and more information from our contacts, in a phenomenon that has been called the disintermediation of news access [31]. While propaganda and political use of fabricated news has always been around, we are now witnessing an explosion of this phenomenon in the current times. Many journals, papers, television shows and naturally social networks devoted a great attention to similar issues [39] as 'fake news', 'post-truth', 'alternative facts' [17] with the division of citizens in various comfort areas of people with similar views [28]. Since the number of people taking social media as their main source of news is steadily increasing worldwide, various narratives, if not directly alternative facts, are circulating in the community without any control from fact-checking institutions. Progress in machine learning and artificial intelligence adds another problem to this already disturbing situation, that is by creating software devoted to the creation and dissemination of fake news. This phenomenon of 'bots' is particularly important on Twitter. Scientists made their best efforts to fight the never ending plague of malicious bots populating social networks. The literature suggest ways to identify them, including studying the profile [20, 38], the network structure [32, 41, 42], and the posting-characteristics [19, 27] of the accounts. Typically the most successful approaches combine studying the topology of contacts (i.e. network analysis) with semantic analysis (the study of the content of the messages) finding a way to remove most of the malicious agents from the system.

A second effect relates to the complexity of financial systems as a result of the increased facility in establishing connections and contracts. Disintermediation in economics and finance indicates another field in which complexity science can help in solving societal problems. One example of these is understanding the interactions between the interconnected layers of the financial system that feed the world economy. Credit to firms, insurances on these contracts (resulting in derivatives of the main loans) has grown exponentially in recent times, given the relative facility with which profit can be done. The typical network interdependence, while improving efficiency, is also related to a new kind of risk. Networks can project the effects of distant problems into our little community. Modern extended enterprises must be aware that if they benefits from this situation in terms of improved efficiency and flexibility, they also have to deal with a number of new risks associated with dependence on distant suppliers. Similarly, financial institutions and governments must be as aware of the risks of this new world as they are aware of the opportunities. These are the salient features of network interconnections—if good things can move easily through a network, then so can bad ones. The very interconnectedness that underpins network efficiency is an Achille's heel when things go wrong. In the simplest version, by modelling financial market as a network, i.e. by representing banks as nodes and contracts as links, it is easy to realise that banks and contracts, (vertices and edges), depend upon the state of the other elements in the same network [9]. Dynamical process of continuous readjustments are then needed in order to ensure the stability of the system. More generally, banks (not to mention the companies they finance) form various multilayer networks defined by the various kinds of contracts with each other in a closed structure where the liabilities of one institutions are the assets of the others [23]. Not surprisingly, the probability that one institution could default depends strongly upon the default probability of all of the other institutions in the network [10]. This is a particularly interesting situation since it gives a possibility to study how to mitigate the financial risk, that is to reduce the probability that a vast majority of the financial institutions might go bankrupt at the same time. A disaster of this kind arises from a variety of different causes, to name a few:

- (i) *An incomplete knowledge of the market*. In the aforementioned cases small errors in the knowledge of the network of contracts can lead to large errors in the probability of systemic defaults [10]. This is due to the amplifying effects related to the multiplicative interplay among the parameters that matter for the default thresholds. Similarly also errors in the contract structure also increase the potential information asymmetries [9]. To make matters worse, the institutions involved may pass on distress both to the institutions that are exposed to them and to the securities they are exposed to [7].
- (ii) *The various kind of contracts that can bind institutions together*. A further complication is related to derivative instruments; they are a special kind of contract between two or more parties where the value of the payoff is derived from the value of another financial instrument or asset (the underlying entity). Originated as insurance, they rapidly changed their purpose. Credit default swap, for instance, are insurances where one party (originally the lender of a loan) buys insurance from a third party to insure against default of another institution (originally the borrower of the loan). However, in a way that contrasts with conventional insurance, in which we own the goods (house, car, contract) we want to protect, credit default swaps can be negotiated on any underlying entity—meaning anyone could take out insurance on our goods. Imagine the ease of mind in leaving your car in a car park when you know that everybody in the neighbourhood can apply to recieve a lump sum if your car is stolen. Speculation therefore emerges as another reason to trade in derivatives, a dubious syndrome that gave rise to 'financial weapons of mass destruction' [15].
- (iii) *The number of banks and* (iv) *the density of connections between them* are a serious issue. It may seem counterintuitive, but observinf the parallels between financial systems and ecosystems [30, 34] we can exploit most of the results of ecosystem stability to describe how risk can propagate in a financial system. In particular, processes that are widely believed to stabilize the financial system, that is, market integration and diversification, can actually drive it towards instability, as they contribute to create cyclical structures which tend to amplify financial distress, thereby undermining systemic stability and making large crises more likely [6]. This effect, which derives from the analysis of the spectral properties of the matrix of leverage, is one among several results that are very hard to discern without the theoretical framework of complex networks.

In summary, complexity science with the mathematical framework of network science represents a crucial instrument to analyse and understand society. While the set of interconnections between us and the institutions operating in this world is constantly growing, new opportunities and new problems arise. I believe that the most formidable challenge, both scientific and societal is to be able to understand the structure of this new highly connected world with simple mathematical instruments. This would allow scientists, policy makers, citizens to make informed decisions and be accountable for their actions. At the same time it is important that complexity science continues to disseminate her results to the general public who may then better understand the multiple features of the world in which they live in.

Acknowledgments

I would like to thank support from European Commission projects HumanE-AI-Net nr. 952026 and SoBigData++ nr. 871042.

No new data were created or analysed in this study.

ORCID iDs

Guido Caldarelli https://orcid.org/0000-0001-9377-3616

References

[1] IBM 2017 *IBM Annual Report* (Endicott, NY: IBM)
[2] Albert R and Barabási A-L 2002 Statistical mechanics of complex networks *Rev. Mod. Phys.* **74** 47
[3] Anderson P W 1972 More is different *Science* **177** 393–6
[4] Brian Arthur W 1999 Complexity and the economy *Science* **284** 107–9
[5] Balassa B 1965 Trade Liberalisation and 'Revealed' Comparative Advantage *The Manchester School* **33** 99–123
[6] Bardoscia M, Battiston S, Caccioli F and Caldarelli G 2017 Pathways towards instability in financial networks *Nat. Commun.* **8** 14416
[7] Battiston S and Caldarelli G 2013 Systemic risk in financial networks *J. Int. Financial Mark. Inst. Money* **1** 129–54
[8] Battiston S, Caldarelli G and Garas A (ed) 2018 *Multiplex and multilevel networks* (Oxford: Oxford University Press)
[9] Battiston S, Caldarelli G, Georg C-P, May R and Stiglitz J 2013 Complex derivatives *Nat. Phys.* **9** 123–5
[10] Battiston S, Caldarelli G, May R, Roukny T and Stiglitz J 2016 The price of complexity in financial networks *Proc. Natl Acad. Sci.* **110** 10031–6
[11] Bianconi G 2018 *Multilayer Networks* vol 1 (Oxford: Oxford University Press)
[12] Boccaletti S, Bianconi G, Criado R, del Genio C I, Gómez-Gardeñes J, Romance M, Sendiña-Nadal I, Wang Z and Zanin M 2014 The structure and dynamics of multilayer networks *Phys. Rep.* **544** 1–122
[13] Boggust A, Carter B and Satyanarayan A 2019 Embedding comparator: visualizing differences in global structure and local neighborhoods via small multiples (arXiv:1912.04853)
[14] Bollobás B 2001 *Random Graphs* (Cambridge: Cambridge University Press)
[15] Buffet W 2002 Berkshire Hathaway's 2002 annual letter *Technical report* (Omaha, NE: Berkshire Hathaway)
[16] Burn-Murdoch J 2012 Study: less than 1% of the world's data is analysed, over 80% is unprotected *Guardian* (https://www.theguardian.com/news/datablog/2012/dec/19/big-data-study-digital-universe-global-volume)
[17] Cadwalladr C 2017 The great British Brexit robbery: how our democracy was hijacked *Guardian* (https://www.theguardian.com/technology/2017/may/07/the-great-british-brexit-robbery-hijacked-democracy)
[18] Caldarelli G 2007 *Scale-Free Networks* (Oxford: Oxford University Press)
[19] Chavoshi N, Hamooni H and Mueen A 2016 DeBot: Twitter bot detection via warped correlation *2016 IEEE 16th Int. Conf. on Data Mining* (Piscataway, NJ: IEEE) pp 817–22
[20] Cresci S, Di Pietro R, Petrocchi M, Spognardi A and Tesconi M 2015 Fame for sale: efficient detection of fake Twitter followers *Decis. Support Syst.* **80** 56–71
[21] Daron J D, Lorenz S, Wolski P, Blamey R C and Jack C 2015 Interpreting climate data visualisations to inform adaptation decisions *Clim. Risk Manage.* **10** 17–26
[22] Darwin C 1859 *On the Origin of Species* (London: Murray)
[23] van Lidth de Jeude J A, Aste T and Caldarelli G 2019 The multilayer structure of corporate networks *New J. Phys.* **21** 025002
[24] Donnat C, Zitnik M, Hallac D and Leskovec J 2018 Learning structural node embeddings via diffusion wavelets *Proc. of the 24th ACM SIGKDD Int. Conf. on Knowledge Discovery & Data Mining* (New York: ACM) pp 1320–9
[25] Dorogovtsev S N and Mendes J F F 2002 Evolution in networks *Adv. Phys.* **51** 1079–187
[26] Gell-Mann M 2002 What is complexity? *Complexity and Industrial Clusters. Contributions to Economics* (Berlin: Springer) pp 13–24
[27] Giatsoglou M, Chatzakou D, Shah N, Beutel A, Faloutsos C and Vakali A 2015 ND-sync: detecting synchronized fraud activities *Advances in Knowledge Discovery and Data Mining* ed T Cao, Lim E-P, Zhou Z-H, Ho T-B, Cheung D and Motoda H (Berlin: Springer) pp 201–14
[28] Grinberg N, Joseph K, Friedland L, Swire-Thompson B and Lazer D 2019 Political science: fake news on Twitter during the 2016 U.S. presidential election *Science* **363** 374–8
[29] Gross T and Sayama H (ed) 2009 *Adaptive Networks* (Berlin: Springer)
[30] Haldane A G and May R M 2011 Systemic risk in banking ecosystems *Nature* **469** 351–5
[31] Hall J 2001 *Online Journalism: A Critical Primer* (London: Pluto Press)
[32] Liu S, Bryan H and Faloutsos C 2017 HoloScope: topology-and-spike aware fraud detection *ACM on Conf. on Information and Knowledge Management, CIKM 2017* (*Singapore, 6–10 November 2017*) (New York: ACM) pp 1539–48
[33] Mantegna R N 1999 Hierarchical structure in financial markets *Eur. Phys. J.* B
[34] May R M and Arinaminpathy N 2010 Systemic risk: the dynamics of model banking systems *J. R. Soc., Interface* **7** 823–38
[35] Mosconi F, Julou T, Desprat N, Sinha D K, Allemand J-F, Croquette V and Bensimon D 2008 Some nonlinear challenges in biology *Nonlinearity* **21** T131–47
[36] Newman M E J 2003 The Structure and function of complex networks *SIAM Rev.* **45** 167–256
[37] Sugihara G, Dixon P A and Milicich M J 2001 Noise and nonlinearity in an ecological system *Nonlinear Dyn. Stat.* 339–64
[38] Satya P R B, Lee K, Lee D, Tran T and Zhang J 2016 Uncovering fake likers in online social networks *25th ACM Int. Conf. on Information and Knowledge Management, CIKM 2016* (*Indianapolis, IN, 24–28 October 2016*) (New York: ACM) pp 2365–70
[39] Tandoc E C, Lim Z W and Ling. R 2018 Defining fake News *Digital Journalism* **6** 137–53
[40] Tumminello M, Aste T, Di Matteo T and Mantegna R N 2005 A tool for filtering information in complex systems *Proc. Natl Acad. Sci. USA* **102** 10421–6

[41] Wang B, Gong N Z and Fu. H 2017 GANG: detecting fraudulent users in online social networks via guilt-by-association on directed graphs *2017 IEEE Int. Conf. on Data Mining (ICDM)* (Piscataway, NJ: IEEE) pp 465–74

[42] Yuan S, Wu X, Li J and Lu A 2017 Spectrum-based deep neural networks for fraud detection *CIKM '17 Proc. of 2017 ACM on Conf. on Information and Knowledge Management (November 2017)* (New York: ACM) pp 2419–22