Structural Patterns in Complex Systems: A network perspective

Christos Ellinas
IDC in Systems
University of Bristol
Bristol, UK
ce12183@bristol.ac.uk

Neil Allan
Systemic Consult Ltd
Bath, UK
mnsnda@management.bath.ac.uk

Anders Johansson
IDC in Systems
University of Bristol
Bristol, UK
A.Johansson@bristol.ac.uk

***Abstract* — Our desire to deliver increased functionality while setting tighter operational and regulative boundaries has fueled a recent influx of highly-coupled systems. Alas, our current capacity to successfully deliver them is, evidently, still in its infancy [1-5]. Understanding how such systems are structured, along with how they compare with their natural counterparts, can play an important role in bettering our capacity to do so. The following article will be grounded upon the principles of network science in order to contrast such naturally evolved systems with systems that we purposefully engineer. Assuming that the underlying structural variety of such systems fuels design uncertainty, and by adopting an evidence-based methodology, systems of the latter class will be compared in terms of their adherence to statistical normality.**

Keywords— risk; project management failures; complex systems; network science; statistical analysis

I. Introduction

The pinnacle of human ingenuity lies in our ability to uncover natural phenomena, understand their underlying drivers and harvest them by engineering purposeful systems [6].We have championed problems found both within the domain of *simplicity* (through the paradigm of reductionism in the 19th century) and *disorganized complexity* (through statistical mechanics in the 20th century) [7]. Alas, our modern society is becoming increasingly reliant upon the delivery of projects characterized by emergent functionality, non-linearity and feedback between tightly coupled, yet fuzzily defined systems. Increased demand, limited resources and tighter schedules push as in transcending the uncharted territory of *organized complexity* [7] where our capability to understand, and consequently control, is continuously challenged.

Such challenges are commonly faced in numerous engineering domains – examples include software development, printed circuit board (PCB) design and construction project management. In an attempt to tackle them, they are routinely divided into a set of sub-problems, each with interfaces and dependencies to the rest. This division represents the human perception to a problem [8]; however, such linear depiction contradicts the inherent complexity of the systems that we desire. As a result, unintended consequences, driven by unwanted emergent properties (e.g. interfacing bugs in software; chaotic oscillation in PCBs and cascading failures in construction projects), are becoming increasingly common [1-5]. Emergence of this sort can frequently lead to significant losses in terms of man-hours, resources and often, human life.

In an attempt to explore such real-life consequences, we will limit ourselves within the analytical framework of complex networks, and by adopting an evidence-based approach, we will employ statistical tools in order to compare and contrast different (sub) classes of complex systems. The derived insight is expected to revolve around two fundamental questions:

- How general, and subsequently, transferable are observations and techniques applied within different classes of complex system?
- Is it possible that the topology and structure of some engineered systems result in an inherently more challenging effort to tame them?

The main objective of this paper is to first, evaluate whether naturally evolved systems have any significant differences between engineered systems, in terms of their structure. Subsequently, we will shift our focus to the latter class of systems, where we will present a holistic methodology in order to map the degree upon which their underlying structure adheres to statistical normality.

The importance of such insight revolves around the confident use of a *mean* value (a direct result of the applicability of a Gaussian distribution) in terms of the statistical dependence, variability and occurrence of the structural blocks of such systems. Such assumptions are commonly encountered in the traditional design of a number of systems, ranging from Mobile Wireless Networks [9] to crucial aspects of the entire economy [10].

In the spirit of empirical falsification [11], two formal hypotheses will be presented and consequently, evaluated:

- Hypothesis 1: Relevant universal characteristics, as observed in natural systems, are equally applicable to engineered systems.
- Hypothesis 2: Regardless of their context, all three engineering sub-classes adhere to a meaningful measure of statistical normality in terms of their underlying structure

The article will be structured as follows; first, a brief introduction to network science, along with relevant findings, will be presented in order to provide the methodological background of the paper. The empirical data sets used will then be introduced, along with a new contribution in terms of network datasets – namely in the form of construction projects. The methodology used to evaluate Hypothesis 1 and 2 will be subsequently presented, along with a review of the results. Discussion and concluding remarks will then follow.

II. COMPLEX NETWORKS

Network science is part of the recent burst of methodologies promoted by the complexity science movement, aimed in understanding complex systems [12] through mathematical *abstraction* in the form of graphs, where components, referred to as *nodes*, interact with each other via *links* [13]. Aided by an unprecedented availability of data and computational power, complex networks have proved to be a unifying paradigm between diverse domains of research [14] where important (and often, *universal*) structural characteristics have been identified [15], some of which are briefly reviewed below.

Namely, it has been commonly assumed that interconnections found within a given system did not significantly deviate from a random distribution [16]. Thus, as they were a residual attribute to intrinsic randomness, connections did not play a significant role in the function of the system. Consequently, their composing topology was assumed to be either completely regular or random. Reference [17] showed that in fact, real world systems tend to lie between these two extremities, having a tendency to be highly clustered (a property of regular systems such as lattices) and yet exhibiting relatively small average path lengths, a characteristic of random graphs. Importantly, the extent of this so-called "small-world" (SW) effect has been consequently shown to be linked with the capacity to efficiently control such system [18]. Through empirical data, further research highlighted a significant deviation from the assumed random distribution to a power-law distribution of connections [19] further reinforcing the need to focus on the interconnectivity between components – the basic underlying principle of holism. Through a shift of focus on the level of aggregation, further research proposed that statistically significant patterns of interconnections (i.e. 3 node sub-graphs, referred to as network motifs) form the basic structural building blocks of complex systems [20]. However, as the link between sub-graph structure and functions is still largely debated [21], a distinction between *structural* and functional subgraphs will be utilized within this paper, where the former will not imply the latter [22]. All three shared organizing principles have been the founding stone of "the new science of networks" [23] and will thus serve as the *de facto* starting point of this article.

The interested reader is encouraged to delve into the excellent reviewing works of [14, 24-27].

III. METHOD

Two fundamentally different classes of systems, referred to as *evolved* and *engineered*, will be contrasted and explored, with an emphasis on the latter. Evolved systems will be defined as a class of systems of which their internal structure is a result of a decentralized, co-evolutionary process. Engineered systems will be defined as the result of a centralized, controlled and nested architectural design process – they will be subsequently sub-divided in terms of their context, namely, Software, PCBs and Construction Projects. Previously published empirical datasets for the majority of systems will be used and expanded upon – see following sub-section.

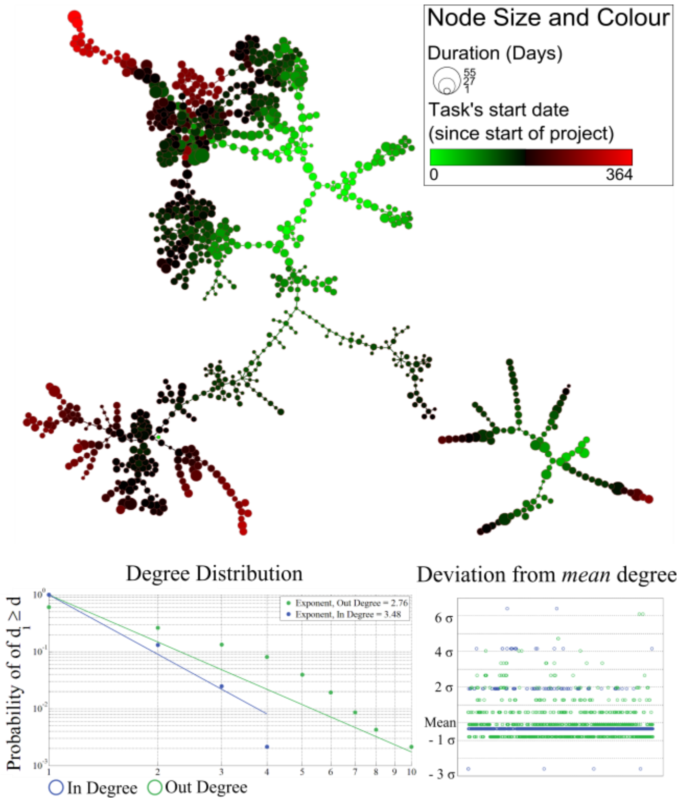

Fig.1. A construction project, mapped as a network via its compromising task dependencies. Bottom left plot illustrates the power law distribution of both in and out degree. Bottom right plot illustrates the deviation from the mean degree - notice the existence of four nodes with a deviation greater than 6 σ (the probability for a node having such great deviation in terms of its out degree is in the order of 5.4×10^{-10}).

A. Data

The entirety of evolved networks and the sub-class of software networks have been attained from literature – see Table 1 for relevant references. Although a limited number of PCB networks were already readily available from Reference [20], further samples were obtained by mapping the relationships between logical gates and inverters for a variety of benchmark circuit, first presented (and consequently, made available) though two international symposia – specifically ISCAS89[37] and ISCAS99 [38].

A notable contribution of this paper is in the form of presenting Construction Projects (abstracted using their compromising *task networks*, explicitly expressed in their relevant Gantt Charts – for an ex. see Fig.1) as a subclass of engineered networks. Note that in some cases, task networks were updated to reflect the on-going progress of the actual project – such data are parenthesized within Table 1 and they have played an equal role in the analysis. Furthermore it is worth noting that in order to limit potential inconsistencies that may arise from endogenous, sociotechnical factors (e.g. organizational culture, internal code of practice etc.) and exogenous (e.g. geopolitical and cultural peculiarities etc.), the Gantt Charts have been obtained from a single source (a regional construction company) in an attempt to minimize the variables that may induce potential inconsistencies.

Special attention was paid in the data harvesting process across both classes in order to ensure that links represent comparable interactions (in this case, functional dependencies).

TABLE 1: DATASETS USED THROUGHOUT THIS PAPER

Class	Node count	Edge count
Engineered - Construction Projects		
S171; (2); (3)	937; (1032); (1093)	1080; (1174); (1200)
S116; (2); (3); (4)	875; (879); (840); (837)	865; (867); (809); (807)
S138; (2); (3); (4)	106; (109); (108); (147)	105; (114); (113); (167)
S107; (2)	520; (522)	561; (564)
S95	184	216
S127	175	194
S132	317	400
S125	730	792
Engineered - Software		
xmms [28]	1032	1096
Digital Material [28]	187	271
MySQL [28]	1501	4212
VTK [28]	788	1375
Abiworld [28]	1096	1830
Linux [28]	5420	11449
Java source code [32]	724	1025
Tulip [34]	111	160
Engineered - PCB		
s208 [20]	122	189
s420 [20]	252	399
s838 [20]	512	819
b11	764	1409
b12	1070	2088
b13	353	611
s1196	561	1027
s1423	749	1238
s1488	667	1387
s9234	5844	8182
s1494	661	1399
s953	440	772
s5378	2993	4391
s713	447	610
Natural		
Email [29]	1133	10903
SW Citations [30]	396	994
Political Blog [31]	1490	19025
Karate Club [33]	63	312
PPI Yeast [35]	1870	4480
Food Web [36]	249	2065
C.Elegans [17]	297	2345

B. *Methodology*

Mathematically speaking, a network can be mapped as a graph G= {{N} {E}} formed by the set N of nodes i ∊ N and the set E of links (i, j) ∊ E, indicating a link from node i to node j (but not necessarily the other way around). An adjacency matrix, **A**, is an aggregated representation of the graph's structure, in which A_{ij}=1 if there is a link between node i and j and 0 otherwise. As the entirety of the datasets is abstracted as directed networks (i.e. links have directionality), A_{ij} is not necessarily equal to A_{ji}, implying the presence of asymmetric adjacency matrices.

1) *First Hypothesis*

In order to evaluate Hypothesis 1, we will first briefly ground our analysis on a coarse level of aggregation. We will utilize a commonly used, statistical measure of a graph, referred to as the graph's diameter (D), and defined as the longest path between a pair of nodes, of which any loops or reuse of a link is forbidden – mathematically defined in equation (1) where eccentricity, ε, is the greatest shortest path between node i and any other node.

$$D = \max_{i \in N} \{\varepsilon(i)\} \tag{1}$$

Similarly, the average path length can be defined as the mean shortest path from node i to j, averaged over all nodes j within the graph – mathematically defined in equation (2) where n is the number of nodes (i.e. the cardinality of set N) and d is the shortest path between i and j.

$$l_{average} = \frac{1}{n}\sum_{j} d_{ij} \tag{2}$$

The relationship between these two metrics can be used to quantify the extent of the "SW" effect – see Figure 3. However, we will need to delve into finer levels of aggregation if we are to embrace and understand the greater extend of structural variety within such systems. To be more precise, we will focus on subgraphs (i.e. a meso level of aggregation) and specifically utilize 3-node subgraphs, as they are often referred to as the structural blocks of complex networks [20, 22, 39]. The freely-available software MAVISTO [40] was employed in order to decompose each system in terms of the 13 possible combinations of 3-node sub-graphs (see Fig.4, inset) and report counts of each one. As subgraph occurrence scales with network size [41], obtained values were then normalized over the total number of subgraphs present, effectively computing values that we will refer to as *subgraph concentration* values – see Fig.2 for results. It is worth noting that the algorithm used allows for the potential reuse of both nodes and links in order to identify a subgraph. This is an important aspect if we are to obtain representative decomposition of each network. By applying a limitation on the potential of reusing either a node or link, significant topological features such as the numerous leaf nodes found in the Construction Projects' networks (as evident in Fig. 1) would not be accounted for.

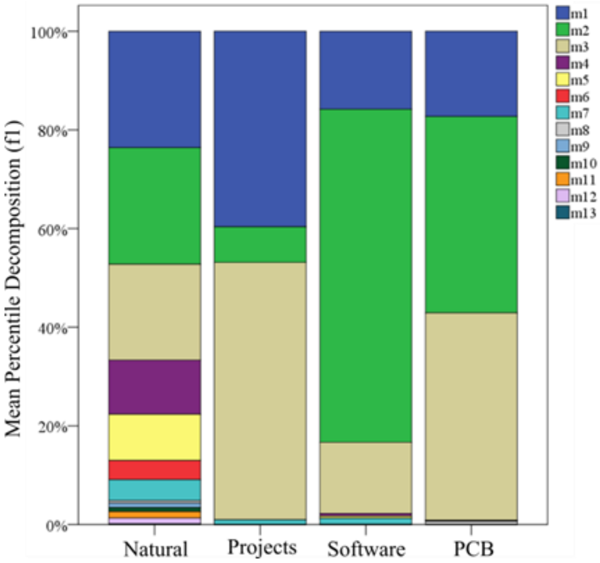

Fig.2. Mean percentile decomposition, in terms of 3-node subgraph concentrations, for both evolved and engineered (the latter is broken down further into the three main three sub-classes). Notice the significantly less variation found within the structure of the latter three.

2) *Second Hypothesis*

Due to the singificant stuctural deviations that are evidently present between the two classes (see Fig. 2), we will remain at the meso-level of detail and further explore how *statistical normality* around the *distribution* of the considered variables (i.e. the four highest subgraph concentrations) within each subclass of the engineered systems varies. QQ plots are used in order to inspect the dispersion between *actual* and *expected* concentrations - the latter were derived from a theoretical Gaussian distribution, computed based on the samples within each subclass. Linear regression (and their R^2 values) will also be used in order to quantify *variation* between all possible (i.e. 6) combinations between the aforementioned subgraph concentrations. Finally, Spearman Correlation coefficients will report any *statistical dependence* between such pairs – note that this coefficient is non-parametric and thus, imposes no assumptions in terms of their underlying distribution.

Effectively, the proposed methodology for testing Hypothesis 2 is grounded on the assumption that the less meaningful an average value is (based on deviation from a Gaussian distribution) the less ordered an engineered system is. Dispersion focuses around the ability to predict the subgraph concentration within each system while both R^2 and Spearman values focus on the confidence of one to say that an observed increase in one structural variable will influence another, assuming that all other variables remain unchanged. Figure 4 summarizes the results.

IV. RESULTS

A. *Macroscopic Network Analysis*

The more pronounced the "SW" effect [16] is within a network, the fewer controllers one may need in order to exert an overall influence upon the entire network [17], thus it is worth exploring its extent within each system. With respect to Fig. 3, left plot, it is important to note that the entirely of systems appears to follow a well-defined linear trend, regardless of the systems' purpose, function, scope or age. One could thus infer that, at this level of aggregation, both classes share a common organizing principle. Notably, the majority of the engineered class appears to dominate the higher region of the plot – PCBs and Construction Projects tend to occupy the higher end whilst Software and Evolved networks are restricted to the lower end. This is mainly due to the acyclic, tree-like structure of the former and implies significant effort to reach (and consequently, manipulate) distant nodes efficiently. By introducing size (in terms of node count) as a variable (see Fig. 3, middle plot) first note that the coherency between the networks now appears to break. Furthermore, we similarly observe that both PCBs and Construction Projects process a much steeper gradient when compared to the rest of the networks, illustrating a greater sensitivity in terms of scalability. It is worth noting that the mean degree for the engineered class appears to be almost scale invariant (Fig. 3, right plot) – this is clearly not the case for the majority of the natural networks.

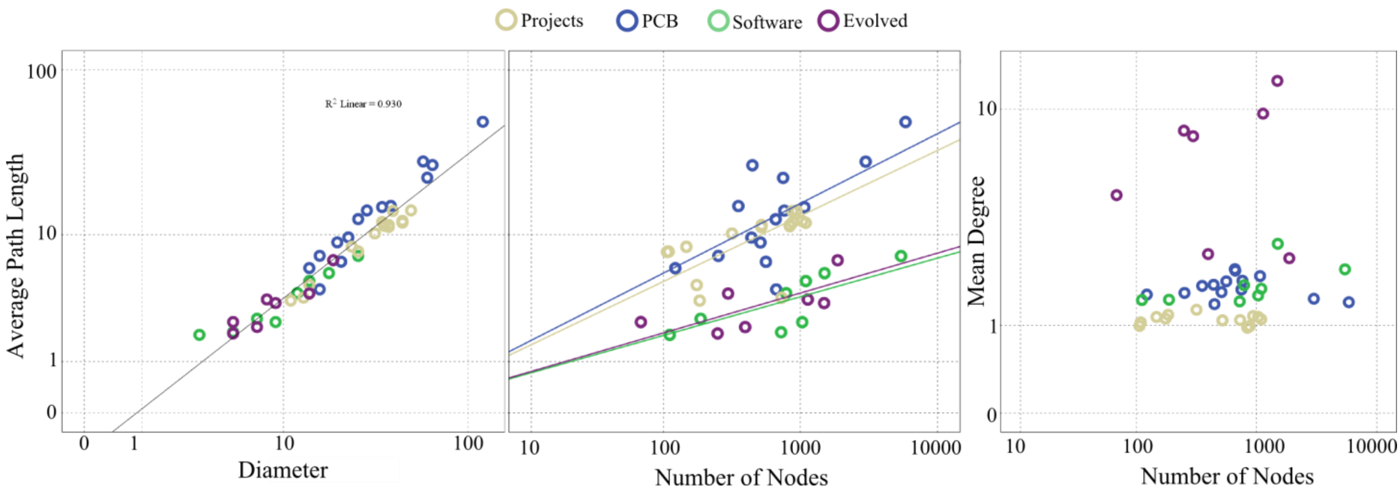

Fig.3. Overall reachability capacity of the network (quantified by the average path length) as function of a network's diameter and number of nodes

Importantly, as engineering systems scale up, the need to exert control must not be limited at a local level as the *average local capacity* to influence the overall network is reduced as systems scale up. This is due to the fact that as the mean degree remains scale invariant, the average path length does not. Within a pragmatic context, such evidence appears to highlight the need to transition from the micro-management of components to a more holistic approach in order to keep up with modern, mega engineering systems. Examples of such counterintuitive insight may include the failure to effectively and efficiently control the destiny of a Construction Project (in terms of timely delivery) by merely micromanaging and optimizing aspects of its constituent, day-to-day tasks.

B. Meso-Scale Network Analysis

Although initial evidence at the macro level may suggest some structural commonalities between the two main classes (Fig. 3, left plot), qualitatively different behavior (Fig. 3, middle and right plot) was also noted. By focusing on the subgraph concentration of the two main classes (as seen in Fig. 2) it is obvious by mere inspection that the Engineering class is significantly less varied when compared to the Evolved class. Consequently, Hypothesis 1 can now be considered to be falsified as evidence at both macro and meso level indicates significant differences between the two classes. Note that Software and Construction Project networks are mainly acyclic (i.e. they do not contain and loops) and thus have access to a limited palette of available subgraphs (namely m1, m2, m3 and m7). On the other hand, PCBs are cyclic and thus have access to all 13 possible combinations. Thus, it is rather remarkable that both Construction Projects and PCBs exhibit relatively similar concentration profiles (in terms of m1, m2 and m3) - see Fig. 2. Nevertheless, the former do illustrate cyclic dependencies (m8), a feature which is not available to the latter. On the other hand, although Software networks draw from the same potential subgraph pool (both are acyclic) they have pronounced differences in the concentration of m1, m2 and m3.

We will now focus on evaluating the second Hypothesis by introducing our suggested methodology, results of which are summarized in Fig. 4. More specifically, we will focus in linking structural variations (in terms of subgraph concentration) within the Engineered class, and thus, rank its three subclasses systems in terms of their predictability, both in terms of subgraph occurrence but also in terms of mapping dependencies between pair of subgraphs.

With reference to Fig. 4, the upper triangular of each matrix plot presents the computed Spearman Correlation coefficients. All three subclasses exhibit a statistically important (i.e. at a 0.01 level) correlation between m1 and m2 and thus, this relationship appears to be of special importance – the authors are currently involved in further exploring the importance of this relationship. Construction Projects further exhibit a significant correlation between m3 and m7 while PCBs similarly exhibit a strong correlation between the pair m2/m3. Thus, in this sense, both Construction Projects and PCBs imply a greater predictability in their internal structure - for ex. an increase in its m2 concentration will fuel expectations of noting a reduction in the m3 concentration within the PCB subclass, assuming all other variables remain constant. By inspection, one can also note that PCBs have the highest *average* R^2 value (each concentration plot can be found on the lower triangular of each matrix), though in absolute terms, they are still relatively low, implying that non-trivial interactions between the subgraph concentrations are at play. Nevertheless, such evidence can serve as proxies for practitioners in terms of project feasibility. For example, a small scale PCB designer would expect a greater success rate when transitioning to larger scale projects than a construction project manager or a software engineer due to the reduced amount of noise found between the interactions of its internal structural blocks.

We now shift our focus on the QQ plots (Fig. 4, diagonal of each matrix plot) where the y-axis represents the expected subgraph concentration (based in a theoretical, normally distributed sample of equal size) of the four most frequent subgraphs per subclass, compared to the observed quantity. Within the Construction Projects' subclass, we observe a relative adherence on the y=x symmetry line, implying a significant convergence between observed and expected value with the notable exception of m2 concentration. Moving on to the Software subclass, we observe similar uniformity with the notable exception of m1, which tends to produce significant deviations. Surprisingly, however, PCBs exhibit the greatest deviation between observed and theoretical values throughout all four subgraph concentrations – this is rather surprisingly as PCBs tend to be more ordered in terms of the expected dependencies between subgraph concentrations, as it was noted above. Increased dispersion between the two quantities can have important implications within a pragmatic context. Knowledge generation from past experience; generic tools and methodological applicability are all examples that fundamentally build on the expectation of what is to be encountered will resemble what has already been encountered and accounted for. However, evidence from this work suggests that this may not always be the case for certain subclasses. Consequently, as actual subgraph concentrations tend to deviate from expected values, the architecture that they represent (and thus, the tools that have been developed to account for their features) will not be applicable to the entire range of apparently equivalent systems.

Finally, the results, as summarized in Fig. 4, can also be used to confidently falsify Hypothesis 2.

V. CONLCUSIONS

This paper has adopted a top-down approach by first comparing two general classes of systems at two levels of aggregations – namely the macro (through an examination for the extent of the "SW" effect [17]) and meso (though the utilization of 3-node subgraphs) level, with a special focus on the latter.

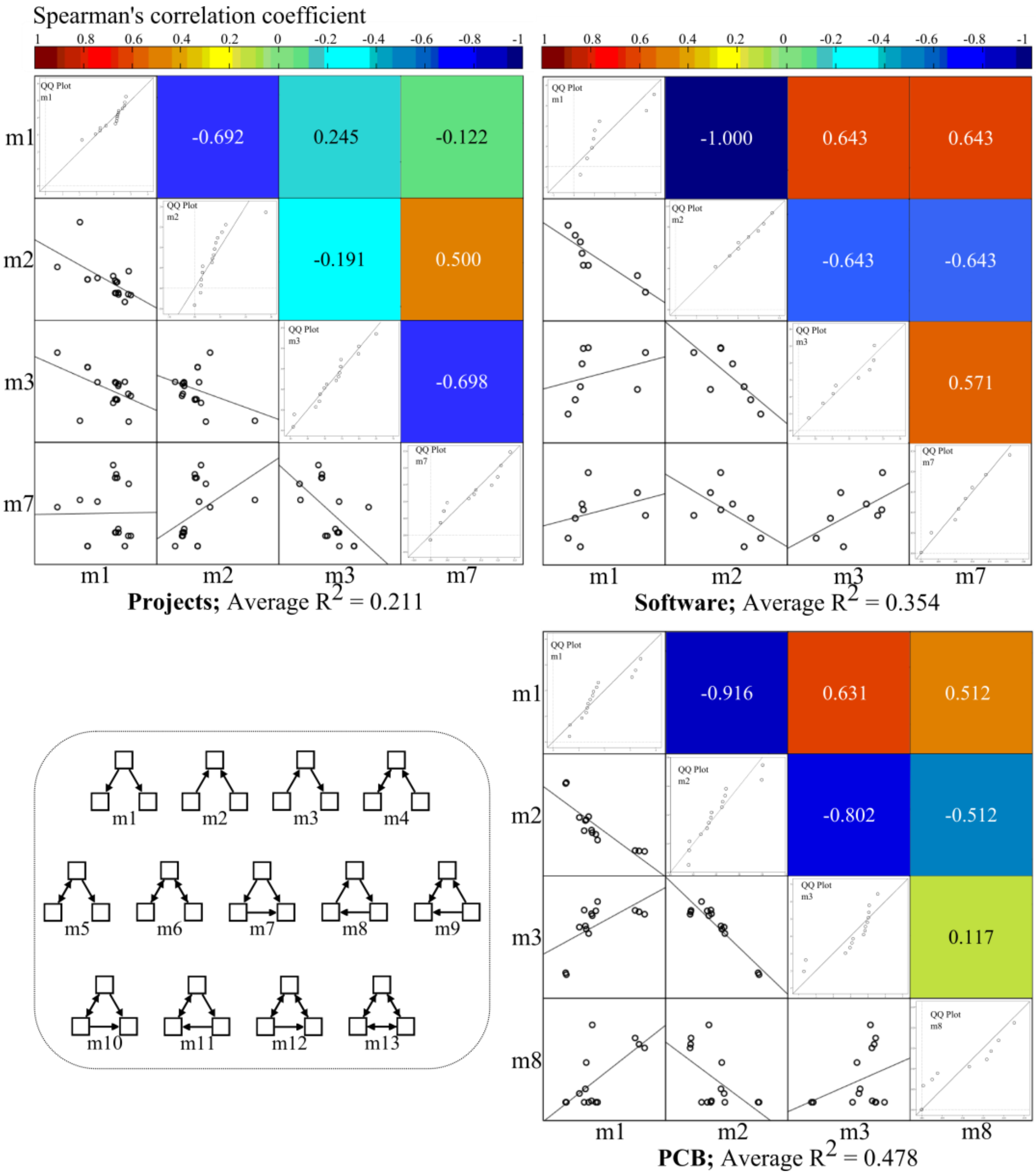

Fig.4. Matrix scatter plots for each of the three engineered subclasses. Upper triangular part of each matrix presents the Spearman Correlation value, ranging from -1 (perfect, negative, statistical dependence) to +1 (perfect, positive, statistical dependence). The lower triangular of each matrix plot presents a linear regression between all possible pairs of the four highest subgraph concentrations. The diagonal illustrates QQ plots where the expected subgraph concentration, based upon a Gaussian distribution, (y-axis) is plotted against the actual subgraph concentration (x-axis). The inset illustrates all possible, 3-node subgraphs.

Focusing on the Engineered class, pronounced subgraph patterns were identified and used in order to falsify the hypothesis that both natural and engineered systems have similar structural decomposition, when their form is abstracted using networks. As such, it poses an important question of whether domain specific knowledge has limited applicability in terms of transferability and applicability within different contexts, both in terms of general insight but also in terms of tool applicability. Furthermore, as topology plays a fundamental role on the dynamic processes that take place both between and within each node [24], along with how such dynamic aspects can eventually feed-back and drive their capacity to adapt [42], such distinct differences may be features that deserve further investigation if we are to engineer systems that mimic desirable aspects of the Evolved class (ex. increased robustness to external perturbations).

While in the process of falsifying the hypothesis that engineered systems possess some sort of meaningful statistical normality, a greater question appears to emerge. Much of our ability to engineer purposeful systems is driven by our capacity to observe patters and deduce theories around their workings – one of the most successful being the Central Limit Theorem, which effectively underlies the importance of any mean value. However, evidence presented within this work, but also within the complexity science BoK, highlight the limited applicability of such assumptions that tend to underlie a number of traditional design methodologies. Examples of such insight present an opportunity for two distinct communities that focus on understanding (complexity science) and delivering such complex systems (engineering) to engage in a constructive dialogue if we are to better our capacity to efficiently deliver such systems.

ACKNOWLEDGMENT

CE acknowledges support from the EPSRC funded IDC in Systems, at the University of Bristol, and Systemic Consult Ltd. Data contribution by LOIS Builders Ltd is gratefully acknowledged. The authors would also like to thank the developers of MAVISTO for making it freely available.

REFERENCES

[1] Williams, T., *Modelling complex projects* 2008: John Wiley & Sons.

[2] ICCPM, *Global Perspectives and the Strategic Agenda to 2025*, in *Complex Project Management* 2011, International Centre for Complex Project Management: Australia.

[3] Hass, K.B., *Managing Complex Projects: A New Model* 2009: Kogan Page, Limited.

[4] DoD, *Competency Standard for Complex Project Managers*, 2006, Commonwealth of Australia: Australia. p. 135.

[5] Bar-Yam, Y. When systems engineering fails-toward complex systems engineering. in Systems, Man and Cybernetics. 2003. IEEE.

[6] Arthur, W.B., The nature of technology: What it is and how it evolves2009: Free Press.

[7] Weaver, W., *Science and complexity.* American scientist, 1948. **36**(4): p. 536-544.

[8] Sterman, J., *Business dynamics* 2000: Irwin-McGraw-Hill.

[9] Tyrakowski, T. and Z. Palka, *A Random Graph Model of Mobile Wireless Networks.* Electronic Notes in Discrete Mathematics, 2005. **22**: p. 311-314.

[10] Mandelbrot, B.B. and R.L. Hudson, The Misbehavior of Markets: A fractal view of financial turbulence2007: Basic books.

[11] Popper, K.R., *Conjectures and refutations*. Vol. 28. 1963: Routledge & Kegan Paul London.

[12] Newman, M..E., *Complex Systems: A Survey.* Am. J. Phys., 2011. **79**(arXiv: 1112.1440): p. 800-810.

[13] Wasserman, S., *Social network analysis: Methods and applications*. Vol. 8. 1994: Cambridge university press.

[14] Newman, M.E., *The structure and function of complex networks.* SIAM review, 2003. **45**(2): p. 167-256.

[15] Barabási, A.-L., *Scale-free networks: a decade and beyond.* Science, 2009. **325**(5939): p. 412-413.

[16] Erdos, P. and A. Rényi, *On the evolution of random graphs.* Publ. Math. Inst. Hungar. Acad. Sci, 1960. **5**: p. 17-61.

[17] Watts, D.J. and S.H. Strogatz, *Collective dynamics of 'small-world'networks.* Nature, 1998. **393**(6684): p. 440-442.

[18] Wang, X.F. and G. Chen, *Synchronization in small-world dynamical networks.* International Journal of Bifurcation and Chaos, 2002. **12**(01): p. 187-192.

[19] Barabási, A.-L. and R. Albert, *Emergence of scaling in random networks.* Science, 1999. **286**(5439): p. 509-512.

[20] Milo, R., et al., *Network motifs: simple building blocks of complex networks.* Science, 2002. **298**(5594): p. 824-827.

[21] Ingram, P., M. Stumpf, and J. Stark, *Network motifs: structure does not determine function.* BMC genomics, 2006. **7**(1): p. 108.

[22] Sporns, O. and R. Kötter, *Motifs in brain networks.* PLoS biology, 2004. **2**(11): p. e369.

[23] Watts, D.J., *The" new" science of networks.* Annual review of sociology, 2004: p. 243-270.

[24] Newman, M.E., *Networks: an introduction* 2009: Oxford University Press.

[25] Dorogovtsev, S.N. and J.F. Mendes, *Evolution of networks.* Advances in physics, 2002. **51**(4): p. 1079-1187.

[26] Boccaletti, S., et al., *Complex networks: Structure and dynamics.* Physics reports, 2006. **424**(4): p. 175-308.

[27] Albert, R. and A.-L. Barabási, *Statistical mechanics of complex networks.* Reviews of modern physics, 2002. **74**(1): p. 47.

[28] Myers, C.R., Software systems as complex networks: Structure, function, and evolvability of software collaboration graphs. Physical Review E, 2003. **68**(4): p. 046116.

[29] Guimera, R., et al., *The real communication network behind the formal chart: Community structure in organizations.* Journal of Economic Behavior & Organization, 2006. **61**(4): p. 653-667.

[30] Batagelj, V. and A. Mrvar, *Pajek datasets.* Web page http://vlado.fmf.uni-lj.si/pub/networks/data, 2006.

[31] Adamic, L.A. and N. Glance. The political blogosphere and the 2004 US election: divided they blog. in Proceedings of the 3rd international workshop on Link discovery. 2005. ACM.

[32] Long, Y. Topology structure and centrality in a java source code. 2012.

[33] Zachary, W.W., *An information flow model for conflict and fission in small groups.* Journal of anthropological research, 1977: p. 452-473.

[34] Auber, D., et al., The tulip 3 framework: A scalable software library for information visualization applications based on relational data. 2012.

[35] Jeong, H., et al., *Lethality and centrality in protein networks.* Nature, 2001. **411**(6833): p. 41-42.

[36] Bascompte, J., C.J. Melián, and E. Sala, *Interaction strength combinations and the overfishing of a marine food web.* Proceedings of the National Academy of Sciences of the United States of America, 2005. **102**(15): p. 5443-5447.

[37] Brglez, F., D. Bryan, and K. Kozminski. Combinational profiles of sequential benchmark circuits. in Circuits and Systems, 1989., IEEE International Symposium on. 1989. IEEE.

[38] Brglez, F. and R. Drechsler. Design of experiments in CAD: Context and new data sets for ISCAS'99. in Circuits and Systems, 1999. ISCAS'99. Proceedings of the 1999 IEEE International Symposium on. 1999. IEEE.

[39] Milo, R., et al., *Superfamilies of evolved and designed networks.* Science, 2004. **303**(5663): p. 1538-1542.

[40] Schreiber, F. and H. Schwobbermeyer, *Towards motif detection in networks: frequency concepts and flexible search.* Proc. Intl. Wsh. Network Tools and Applications in Biology (NETTAB'04), 2004: p. 91-102.

[41] Valverde, S. and R.V. Solé, Network motifs in computational graphs: A case study in software architecture. Physical Review E, 2005. **72**(2): p. 026107.

[42] Gross, T. and B. Blasius, *Adaptive coevolutionary networks: a review.* Journal of the Royal Society Interface, 2008. **5**(20): p. 259-271.