Ellinas, C., Allan, N. D., & Johansson, A. F. (2016). Exploring Structural Patterns Across Evolved and Designed Systems: A Network Perspective. *Systems Engineering*, *19*(3), 179-192. https://doi.org/10.1002/sys.21350

## University of Bristol - Explore Bristol Research

### General rights

# Exploring structural patterns across Evolved and Designed Systems: A network perspective

Christos Ellinas[1, 2, #], Neil Allan[2,*], Dr. Anders Johansson[1, &]

[1]IDC in Systems, University of Bristol, Bristol, UK,
[2]Systemic Consult Ltd, Bradford-on-Avon, UK
[#]ce12183@bristol.ac.uk; Telephone Number: +44 (0) 1225 282251
[*]neil@systemicconsult.com; Telephone Number: +44 (0) 1225 282251
[&]A.Johansson@bristol.ac.uk; Telephone Number: +44 (0) 117 954 5110

***Abstract* — Our desire to deliver increased functionality while setting tighter operational and regulative boundaries has fueled a recent influx of highly-coupled systems. Nonetheless, our current capacity to successfully deliver them is still in its infancy. Understanding how such Designed systems are structured, along with how they compare to their naturally Evolved counterparts, can play an important role in bettering our capacity to do so. Based on this premise, the structural patterns underlying a wide range of seemingly unrelated systems is uncovered using tools from network science. By doing so, structural patterns emerge and are subsequently used to highlight both similarities and differences between the two classes of systems. With a focus on the Designed class, and assuming that increased structural variety fuels design uncertainty, it is shown that their adherence to statistical normality (i.e. expected vs. encountered patterns and statistical correlations between combinations of such patterns) is rather limited. Insight of this sort has both theoretical (context agnostic approaches are increasingly relevant within the domain of Systems Engineering, yet are rarely used) and practical (transferability of knowledge) implications.**

***Keywords— Complexity Science, Systems Science, Complex Networks,***

## 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 [Arthur, 2009]. 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) [Weaver, 1948]. Alas, modern society is increasingly faced with challenges driven by undesirable emergence (e.g. the nature of interconnectivity of various sociotechnical systems challenges our ability to contain epidemics across them [Pastor-Satorras and Vespignani, 2001, Vespignani, 2012]); non-linearity (e.g. a single failure can trigger cascades capable of significantly impacting the entire system [Lorenz, Battiston and Schweitzer, 2009]) and limited observability [Liu, Slotine and Barabási, 2013] force us into transcending the uncharted territory of *organized complexity* [Weaver, 1948] where our capability to understand, and consequently control, is eventually 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 [Sterman, 2000]; 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 [Ma, He and Du, 2005]; chaotic oscillation in PCBs [Magistris, Bernardo and Petrarca, 2013] and cascading failures in construction projects [Ellinas, Allan, Durugbo and Johansson, 2015]), are becoming increasingly common [Bar-Yam, 2003, ICCPM, 2011, Punter, 2013, Venkatasubramanian, 2011, Vespignani, 2010, Williams, 2002]. Emergence of such unintended behavior can frequently lead to significant losses in terms of man-hours, resources and often, human life.

Mounting evidence have highlighted the link between such events and the complex nature of the systems that sustain them – partly due to the non-trivial nature of a system's architecture (or topology) [May, Levin and Sugihara, 2008, Schweitzer, et al., 2009, Haldane and May, 2011, Helbing, 2013]. Network science has proven to be increasingly successful in exploring the relationship between the two (emergence of unintended behavior and topology) [Barabási, 2011] yet its adoption by the Systems Engineering community has been relatively constrained [Bellamy and Basole, 2013, Sheard and Mostashari, 2009]. Contributing to its diffusion, this comparative study draws notions from network science to compare and contrast a large set of empirical systems. The focus of this work devolves 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 Designed systems result in an inherently more challenging effort to tame them?

The objective of this paper is to compare the underlying structure of a wide range of systems, from a context agnostic point of view. Typically, a large set of the systems considered (specifically, the Designed class of systems set – see Section III for a definition) are analyzed from a context dependent[1] point of view – which is typical within the engineering domain [Bar-Yam, 1997, Ottino, 2003, Ottino, 2004]. Such view is necessary to develop operational tools yet has noted weaknesses in identifying principles that may be shared across a wide range of

[1] As an example, cconsider the case of Projects, where the dominating view is that each project is "unique endeavours" due to the unique contextual features that characterise its development [PMI, 2008].

systems, with recent attempts adopting a complex systems view to counter this weakness (e.g. [Ellinas, Allan and Johansson, 2016]). Once patterns of similarity are highlighted across seemingly different systems within the Designed class, a methodology is developed to evaluate the degree upon which their structure adheres to statistical normality. By statistical normality, we mean the compounding effect of (a) the difference between actual and expected concentration of dominant patterns of interconnections that describe its structure, and (b), statistical correlations between these patterns. From a theoretical point of view, the occurrence of common patterns across such systems indicates the utility of adopting context agnostic perspectives in identifying unifying themes across Designed systems – a view consistent with the recent work of [Heydari and Dalili, 2015, Polacek, Gianetto, Khashanah and Verma, 2012, Sheard and Mostashari, 2009]. From a practical point of view, the derived insight is relevant to the transferability of knowledge (which assumes that what has been encountered before can be used as a reasonable precursor for what will be encountered next i.e. variability in terms of the structural blocks of a system is limited). Such assumptions are commonly encountered in the traditional design of a number of systems, ranging from Mobile Wireless Networks [Paxson and Floyd, 1995, Tyrakowski and Palka, 2005] to crucial aspects of the entire economy [Ellinas, Allan and Cantle, 2015, Mandelbrot and Hudson, 2007].

To do so, two null hypothesis are introduced:

- Hypothesis 1: Relevant universal characteristics, as observed in evolved systems, are equally applicable to Designed systems.
- Hypothesis 2: Regardless of their context, Designed systems adhere to a statistical normality in terms of their underlying structure

The article is structured as follows: a brief overview on the motivation and relevant literature is first presented, followed by the empirical dataset and the method used to assess the Hypotheses. The results of the analysis are subsequently presented, followed by a discussion and concluding remarks. An Appendix is also included for supporting information.

## II. THEORETICAL BAKCGROUND

### A. Motivation

Complexity is a theme often faced by Systems Engineers [Polacek, Gianetto, Khashanah and Verma, 2012, Sheard and Mostashari, 2009, Bar-Yam, 2003] – but what is it exactly? Dictionary definitions of *complex* include “consisting of interconnected or interwoven parts” and “not easy to understand or analyze” – one can intuitively appreciate how the former leads to the latter, as one needs to describe the parts *and* how each part relates with all the rest to fully describe the state of a complex system [Weaver, 1948, Anderson, 1972, Geli-Mann, 1994, Barabási, 2007]. Due to the wide abundance of complex systems [Newman, 2011] complexity science often adopts context agnostic tools in an attempt to identify widely applicable (and often *universal*) features across a wide range of domains. Despite their recent success in doing so (e.g. [May, Levin and Sugihara, 2008, Watts and Strogatz, 1998, Barzel and Barabási, 2013, Barabási and Albert, 1999]) such approaches have been criticized on their operational limitation, precisely due to their context agnostic nature [Willinger, et al., 2002, Borgatti, Mehra, Brass and Labianca, 2009].

Traditional engineering adopts a markedly different approach – due to its heavily operational nature, there is a distinct focus on specificity [Bar-Yam, 1997] i.e. based on identification and inclusion of contextual variables that influence the behavior of the system (e.g. in the context of Systems Engineering, consider the methodology underlying the construction of a Causal Loop Diagram [Sterman, 2000] or the use of the Quality Function Deployment methodology [Chan and Wu, 2002]). Such approach is necessary due to nature of the engineering method i.e. the explicit focus on meeting a given set of requirements whilst satisfying a set of restrictions – both being a function of the systems’ context [INCOSE, 2015, Koen, 1985]. The surge of technological[2] innovation is a testament on the success of engineering [Arthur, 2009], yet this persistence with system specificity is not without its limitations. Recent concerns have being voiced around the ability of Engineering in general [Ottino, 2003, Ottino, 2004], and Systems Engineering in particular [Bar-Yam, 2003, Sheard and Mostashari, 2009], to cope with the design of complex systems. Indicative, Sheard and Mostashari [2009] note that “most systems engineers do not realize that the systems engineering process […] can be studied by complex systems methods” – contextual abstraction being an integral part of the latter (i.e. complex systems method) [Bar-Yam, 1997].

To contribute to the capacity of Systems Engineering in the successful development of Designed complex systems, this work adopts a complexity view by leveraging Network Science tools [Newman, 2009] in order to compare a wide range of systems. In doing so, relevant insight which can guide the design process underlying complex systems are to be derived. To do so, a context agnostic perspective is imposed across a wide range of system, focusing on highlighting both similarities and differences. We emphasize that a strict engineering perspective (i.e. context dependent[3]) would inevitably lead to increasingly noisy results, where no coherent patters could emerge – a result of the highly dissimilar nature of the dataset, both between and within the two system classes.

### B. Complex Networks

The study of complexity has recently been spearheaded by network science – an interdisciplinary domain grounded on principles of statistical physics, graph theory and computer science and a focus on real-world, of socio-technical systems

[2] The reference to “technology” is used in its general sense to include artefacts, processes and/or methodologies [Arthur, 2009].

[3] In the context of Systems Engineering, typical examples of such context dependent approaches include NAF, DODAF, MODAL and Zachman [Godfrey, 2013].

[Vespignani, 2012, Newman, 2011, Barabási, 2007, Mitchell, 2006].

To do so, systems are abstracted as networks (or graphs), where components, referred to as *nodes*, interact with each other via *links* [Bellamy and Basole, 2013, Basole, et al., 2011]. Aided by an unprecedented availability of data and computational power [Lazer, et al., 2009], complex networks has provided a unifying ground for identifying a number of important topological principles that describe the structure of a wide range of systems [Barabási, 2009, Newman, 2003], some of which are briefly reviewed below. For a deeper exposition of the area, the interested reader is referred to the excellent reviewing work of [Albert and Barabási, 2002, Boccaletti, et al., 2006, Dorogovtsev and Mendes, 2002, Newman, 2003].

It has been commonly assumed that interconnections found within a given system did not significantly deviate from a random distribution [Erdos and Rényi, 1960]. Thus, they could be regarded as a residual attribute of intrinsic randomness and consequently be regarded as irrelevant to the function of the system (with important implications to their design – see [Ottino, 2003]). Seminal work by Watts and Strogatz [1998] showed that in fact, real world systems were balancing between order and randomness [Strogatz, 2001], with 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). Sparking a surge of work around the area (which subsequently gave birth to what is now called “Network Science” [Watts, 2004] ) the ubiquity and importance of this so-called “small-world” (SW) has been illustrate across numerous important processes including diffusion [Karsai, et al., 2011], collective action [Centola, Eguíluz and Macy, 2007] and synchronization [Latora and Marchiori, 2001, Wang and Chen, 2002] – note that all mentioned processes are of relevance to the systems included in this study.

Subsequent work by Milo, et al. [2002] shifted the analysis focus from global measures (such as the average clustering coefficient and average shortest path, as used in [Watts and Strogatz, 1998]) to interconnectivity patterns revolving around 3 nodes. Such patterns are widely referred to as “network motifs” and have been found to be statistically over (or under) represented in real world systems. Being the minimal[4] form of pattern that can capture non-trivial features, network motifs are often referred to as the basic structural building blocks of complex systems [Milo, et al., 2004, Milo, et al., 2002]. It should be noted that even through attempts have been made to network motifs with network function [Alon, 2007], this relationship is largely debated [Ingram, Stumpf and Stark, 2006, Knabe, Nehaniv and Schilstra, 2008]. Hence, a distinction between *structural* and *functional* subgraphs will be adopted within this paper, where the former will *not* imply the latter [Sporns and Kötter, 2004].

## III. METHOD

In the spirit of Milo, et al. [2004], the empirical dataset is divided based on the process that led to the systems’ design, giving rise to two main classes – Evolved and *Designed*. Evolved systems will be defined as a class of systems of which their internal structure is a result of a decentralized, co-evolutionary process. Designed 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. In other words, the classification between Evolved and Designed systems is a function of the process that led to their resulting form rather than the function that they are mean to fulfil (e.g. redistributed traffic control is in effect decentralized [Lämmer and Helbing, 2010], yet it is the product of a top-down, centralized and structured design effort, falling under the Designed class of systems).

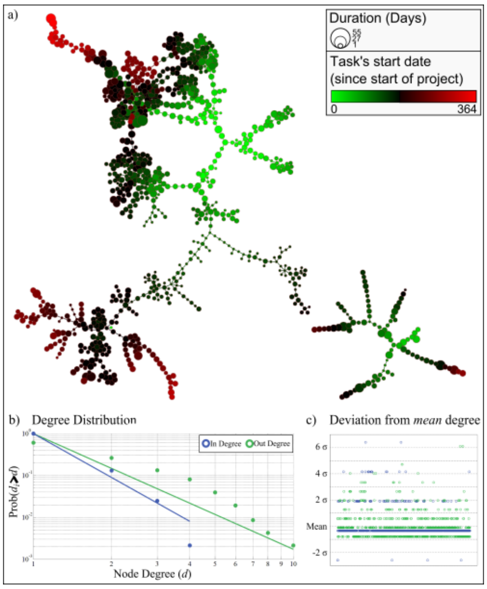

Fig.1. a) A construction project, mapped as a network via its compromising task dependencies; b) plot illustarting the heavy-tail nature of both in and out degree distributions; c) deviation from the mean degree, in terms of standard deviations (z-score). With respect to node out-degree, note the occurance of of four values greater than 6 σ. The probability for one such occurance is in the order of 5.4 x $10^{-10}$.

### A. Data

To ensure compatibility across systems of different domains, special care has been taken to ensure that the network abstractions are comparable i.e. nodes are functional components, with links represent functional dependencies

[4] Theoretically, a dyad would be the minimal network structure, yet it’s capacity to contain useful information around topological features is extremely limited due to the limited possible states that it can offer (i.e. two nodes can either be connected or not).

The entirety of Evolved networks and the sub-class of software networks have been attained from literature – see Table 1. Although a limited number of PCB networks were already readily available from [Milo, et al., 2002], further samples were obtained using the same methodology i.e. by mapping the relationships between logical gates and inverters for a variety of benchmark circuit, first presented (and consequently, made available) through two international symposia – specifically ISCAS89 [Brglez, Bryan and Kozminski, 1989] and ISCAS99 [Brglez and Drechsler, 1999].

Table 1: Datasets used throughout this paper

| Class | Node count | Edge count |
|---|---|---|
| ***Designed - Construction Projects*** | | |
| Project A (1); (2); (3) | 935; (1037); (1093) | 1070; (1198); (1200) |
| Project B (1); (2); (3); (4) | 875; (879); (833); (802) | 865; (867); (806); (810) |
| Project C (1); (2); (3); (4) | 106; (109); (108); (147) | 105; (114); (113); (167) |
| Project D (1); (2) | 521; (514) | 563; (563) |
| Project E | 184 | 216 |
| Project F | 175 | 194 |
| Project G | 312 | 369 |
| Project H | 730 | 792 |
| ***Designed - Software*** | | |
| xmms [Myers, 2003] | 1032 | 1096 |
| Digital Material [Myers, 2003] | 187 | 271 |
| MySQL [Myers, 2003] | 1501 | 4212 |
| VTK [Myers, 2003] | 788 | 1375 |
| AbiWord [Myers, 2003] | 1096 | 1830 |
| Linux [Myers, 2003] | 5420 | 11449 |
| Java source code [Ying and Ding, 2012] | 724 | 1025 |
| Tulip 3 [Auber, et al., 2012] | 111 | 160 |
| ***Designed - PCB*** | | |
| s208 [Milo, et al., 2002] | 122 | 189 |
| s420 [Milo, et al., 2002] | 252 | 399 |
| s838 [Milo, et al., 2002] | 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 |
| ***Evolved*** | | |
| Email [Guimera, et al., 2006] | 1133 | 10903 |
| SW Citations [Batagelj and Mrvar, 2006] | 396 | 994 |
| Political Blog [Adamic and Glance, 2005] | 1490 | 19025 |
| Karate Club [Zachary, 1977] | 63 | 312 |
| PPI Yeast [Jeong, Mason, Barabási and Oltvai, 2001] | 1870 | 4480 |
| Food Web [Bascompte, Melián and Sala, 2005] | 249 | 2065 |
| C.Elegans [Watts and Strogatz, 1998] | 297 | 2345 |

A notable contribution of this paper is the explicit consideration of Projects as complex systems[5], referred to as *task networks*[6] (see Fig. 1). This view that has occasionally surfaced within the Systems Engineering community [Sharon, de Weck and Dori, 2011, Vidal and Marle, 2008] yet was restricted to a conceptual level. To provide for a representative type of project, Construction Project have been used throughout – such projects are a core example of engineering projects, the latter forming the majority of modern organizational activity [Shenhar, 2001]. Specifically, the task network is a view capturing the technological aspect of a project [Baccarini, 1996] where each node is a task, with every functional dependency being represented as a link (often referred to as activity-on-node notation [Valls and Lino, 2001]). 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. As a typical example, consider Project A, which corresponds to an educational institution, with an agreed cost of approximately 15 million USD. The 1st time-shot was produced 40 days after the project was launched – its respective task network is composed of 935 nodes and 1,070 links. The 2nd time-shot was produced 104 days after the project was started, with its respective task network corresponding to 1,037 nodes and 1,198 links. Finally, the 3rd time-shot was produced 212 days after the project was started, with its respective task network corresponding to 1,039 tasks and 1,093 links. Note that in an attempt 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.), project data have been obtained from a single source.

[5] Work around projects typically adopts a process-oriented view e.g. [Vanhoucke, 2013]

[6] For details on the methodology of extracting a task network from a Gantt Chart, see [Ellinas, Allan, Durugbo and Johansson, 2015].

## B. Methodology

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

### 1) Hypothesis 1

In order to evaluate Hypothesis 1, a coarser level of analysis is first adopted by examining the correlation between a graph's diameter ($D$) and average path length ($l_{average}$). This relationship can be interpreted as capturing the overall reachability of a network, and to an extent exemplifies the "SW" effect (see Fig. 3).

Specifically, $D$ is 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, $\varepsilon$, is the greatest shortest path between node $i$ and any other node.

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

Similarly, $l_{average}$ is 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 $j$.

$$l_{average} = \frac{1}{n} \sum_j d_{ij} \qquad (2)$$

Delving further into the analysis, a less aggregated mode of analysis is adopted by focusing at the meso level of the network. Consider the basic building elements of a network i.e. nodes and links, with their resulting combinations (so-called subgraphs) giving rise to the overall network topology. In terms of scale, subgraphs can range from a dyad up to the size of the largest connected component, with larger subgraphs capturing an increased amount of information with respect to possible topological features. Yet such increase in subgraph size significantly increases the complexity of the analysis by increasing the number of all possible combinations that need to be examined (e.g. 13 possible 3-node subgraphs; 201 possible 4-node subgraphs etc.). Deemed to strike a satisfactory balance between adequate variety of combinations and analytical tractability, 3-node subgraphs have been used in several influential studies [Milo, et al., 2002, Milo, et al., 2004, Sporns and Kötter, 2004, Alon, 2007], earning them the moniker of serving as building blocks for complex networks [Milo, et al., 2002]. It is worth noting that their use strikes a balance between local (using locally derived information to characterize the network e.g. degree) and global (using globally derived information to characterize the network e.g. average path length) levels analysis. This is because the use of 3-node subgraphs incorporates information of a node's neighborhood (capturing basic information such as transitivity [Wasserman, 1994]) while being robust to global network features (e.g. bottle-neck nodes [Newman, 2009]) that can skew global measures. The breakdown of convergence in Figure 3b may be one such example, and provides the motivation for shifting the analysis to the meso level.

The freely-available software MAVISTO [Schreiber and Schwobbermeyer, 2004, Schreiber and Schwöbbermeyer, 2005] was employed in order to decompose each system in terms of the 13 possible combinations of 3-node sub-graphs and report counts of each one – see Fig. 2. As subgraph occurrence scales with network size [Valverde and Solé, 2005], obtained values were then normalized over the total number of subgraphs present, effectively computing values that we will refer to as *subgraph concentration* values. 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.

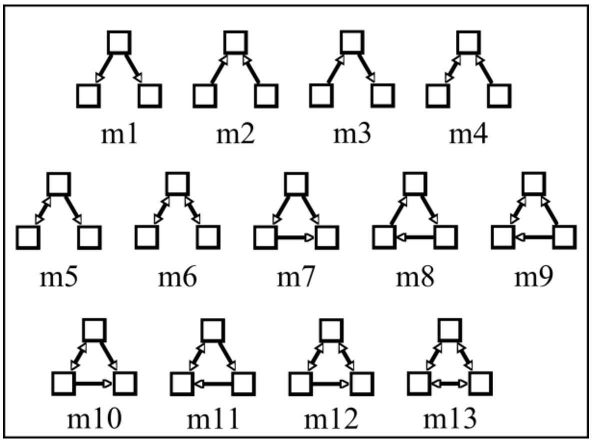

Fig. 2. All 13 possible 3-node subgraphs.

Recent seminal work by Liu, Slotine and Barabási [2011] has introduced the notion of structural controllability, in an attempt to assess the inherent capacity of a network to be controlled. Based on this work, certain subgraphs can be considered to be theoretically (un) controllable, depending on their underlying structure. In particular, the rank of controllability matrix **C,** and whether it matches the number of nodes contained within the subgraph, determines the capacity of a subgraph to be controlled – for further details see Appendix. In the case of 3-node subgraphs, m3 is an example which satisfies this conditions (i.e. $\text{rank}\ \mathrm{C} = N$), while m2 is an example which does not satisfy this condition. As such, Evolved and Designed systems may be further compared from a controllability point of view.

### 2) Hypothesis 2

Focusing at the meso level of analysis, the adherence to statistical normality of the systems will be explored in terms of subgraph occurrence. It is worth noting that evaluating whether a feature is randomly distributed does not necessarily imply that the process in which the system has been developed is random[7]. Rather, it can be used to identify whether a given feature can be explained as the result of pure randomness [Barabási, 2009]. This approach is widely adopted in the study of complex systems in general [Ellinas, Allan and Johansson, 2016, Heydari and Dalili, 2015, Barabási and Albert, 1999, Watts and Strogatz, 1998], and of subgraph occurrence in particular [Milo, et al., 2004, Milo, et al., 2002], where empirical measurements are compared to randomly-distributed equivalent in order to assess whether the two converge.

By focusing on the four highest occurring subgraphs, QQ plots will be used to inspect the dispersion between actual and expected subgraph concentrations – the latter being derived based on the average occurrence of the same subgraph within its respective (sub) class, assuming it is normally distributed. The statistical correlation between subgraph concentrations of all possible combinations of the four main subgraphs is also examined using scatter plots and quantified using the Spearman Correlation coefficients – note that this is a non-parametric measure and thus, imposes no assumptions in terms of the underlying distribution.

Hypothesis 2 is grounded on the expectation that Designed systems will adhere to statistical normality as they have been the result of a controlled, bottom down and centralized design effort. By focusing on the dispersion between actual and expected subgraph concentration, one can test this hypothesis from the point of view of being able to predict the encountered concentrations. In addition, the hypothesis can be tested from a correlation point of view by focusing on the confidence at which one infer an increase in a given structural motif will influence another, assuming that all other variables remain unchanged. Results are presented in Figure 6.

## IV. Results and Discussion

### A. Hypothesis 1

#### 1) Macro-Level Analysis

The "SW" effect influences a substantial number of processes which are of relevance to the entirety of systems examined (see Section II, B) – Fig. 3a quantifies the effect by considering the ratio between average path length and diameter. Specifically, all systems 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 adhere to a common organizing principle. Notably, the majority of the Designed 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. In other words, Software and Evolved systems are in effect "smaller" than PCBs and Construction Project. This is mainly due to the acyclic, tree-like structure of the latter (PCBs and Construction Projects) and implies a significant effort to reach (and consequently, manipulate) distant nodes efficiently.

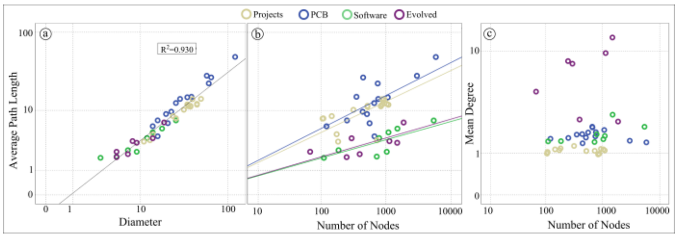

Fig.3. Overall reachability capacity of the network (quantified by the average path length – y-axis) as function of: a) network diameter and b) number of nodes; c) plot of the mean degree (y-axis) against the number of nodes (x-axis).

Such differentiating behavior has important implications on the inherent capacity of a system to exhibit collective behavior. In the case of simple cascades (i.e. a single node is capable of influencing the state of a neighboring node) "smaller" networks can benefit from a higher rate of progression and hence enhance the capacity to be synchronized. On the other hand, such "smaller" networks are increasingly robust to complex cascades (i.e. multiple connected nodes are required to influence the state of a neighboring node). Assuming that the context of the system will dictate the distinct process that is exposed to (i.e. simple or complex cascade) designers may use this insight to define the space in which their system must lie by suitably architecting the underlying structure of that system.

Interestingly, the consistency of the relationship noted in Fig. 3a begins to break once the size of the system is taken into account – see Fig. 3b and 3c. With respect to the relationship between average path length and number of nodes, both PCBs and Construction Projects are defined by a steeper gradient with respect to the rest of the networks, which translates to increased sensitivity in terms of system scalability. By shifting focus to the relationship between average degree and number of nodes (Fig. 3c) a tendency for scale invariability is noted across the class of Designed systems –this is not the case for the majority of the natural networks. By combining the insight of Fig.3b and 3c, consider the case where the capacity to influence the behavior of the system is a function of the ability to reach each node i.e. its average local capacity. As the mean degree of Designed systems remains scale invariant, the ability to efficiently reach distant nodes reduces – a result of the increasing trend of the average path length. In effect, this insight highlights the need to transition

[7] A classic debate on the origin of a feature deviating from statistical normality is the one between Barabási and Albert [1999] and Carlson and Doyle [2002]. Specifically, both studies effectively assume that a feature (number of connections) of a designed system (Internet) is normally distributed, before falsifying it using empirical data. Both studies proceed to provide fundamentally different views in explaining why the system deviates from statistical normality, with Barabási and Albert [1999] proposing an effectively random process (BA model), whereas Carlson and Doyle [2002] argue that this feature reflects purposeful, engineered action (HOT model).

from the micro-management of components to a more holistic approach in order to keep up with the design of large-scale systems. Examples of such counterintuitive insight may include the failure to effectively and efficiently control the progress of a Construction Project by merely micromanaging and optimizing aspects of its constituent, day-to-day tasks.

### 2) *Meso-Scale Analysis*

Focusing on the 3-node subgraph concentrations, the consistency of structural features noted at the macro level breaks down, with the Designed class being significantly less varied than the Evolved – see Fig. 4.

Note that Software and Construction Project networks are mainly acyclic (i.e. they do not contain any loops) and thus have access to a limited palette of subgraphs (namely m1, m2, m3 and m7). Conversely, 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), even though the 4th most frequent subgraph is m8 – a subgraph which is not available to the Construction Project sub-class. Finally, note that even though Software networks draw from the same potential subgraph pool as Construction Projects (both are acyclic) they have pronounced differences in the concentration of m1, m2 and m3.

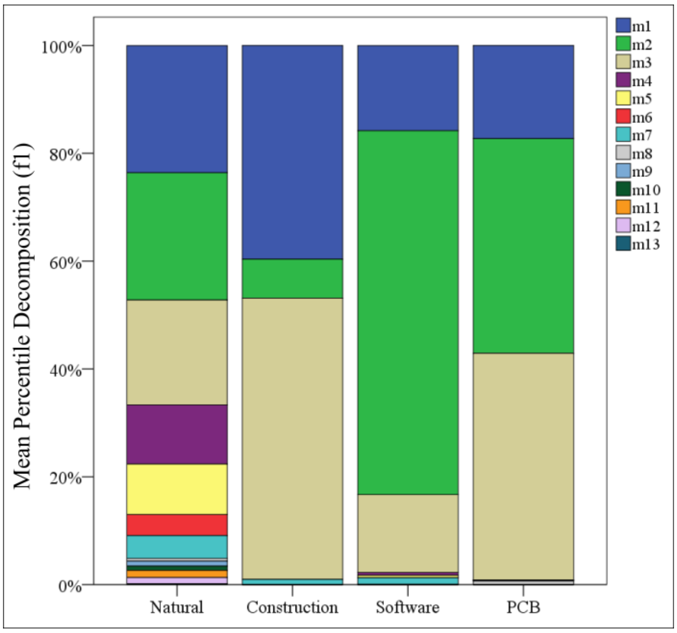

Fig.4. Mean percentile decomposition, in terms of 3-node subgraph concentrations, for both Evolved and Designed. The latter is broken down further into the three main three sub-classes. Notice the significantly less variation found in the structure of the Designed class, compared to the Evolved class

From a controllability point of view, Fig.5 plots the concentration of m2 (theoretically uncontrollable) against m3 (theoretically controllable), with each class having a distinct behaviour. Evidently, the Evolved class is defined by a balancing effect, where an increase in m2 is matched by an increase in m3. Interestingly, this relationship is well approximated by the y=x line, indicating that the effect of introducing more control is counteracted by an increase in the concentration of subgraphs that cannot be controlled. On the other hand, the Designed class is described by a decreasing trend, where an increase in the m3 concentration is matched by a decrease in m2.The contrasting nature of the two classes leads to an interesting question: is this trend the signature of purposeful, human effort in taming (controlling) a complex system? The authors expects that further work around the underlying mechanism that fuels this behaviour could provide further evidence on how these two classes differ, bridging the gap between the two. Nonetheless, such work is beyond the scope of this paper.

In summary, and despite results at the macro-level (Fig.3), analysis at the meso level (Fig. 4 and 5) indicate that the two (sub) classes are indeed distinct, falsifying Hypothesis 1. At the same time, it illustrates the utility of adopting a context agnostic perspective, having the potential to uncover robust patterns, with Fig.5 being a principal example.

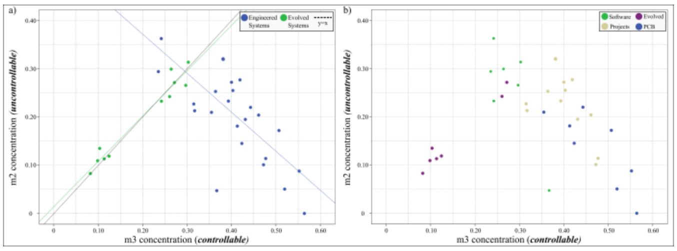

Fig.5. a) Plot of m2 subgraph concentration (y-axis) against m3 (x-axis), the former being theoretically uncontrollable with the latter being controllable. Note the close alignment of the Evolved systems to the y=x line. The converse behavior is shown by the Designed systems, with m1 decreasing as m2 increases; b) same plot with specific sub-classes of the Designed class being shown.

## B. *Hypothesis 2*

Focusing on Hypothesis 2, results summarized in Fig. 6 are used to evaluate the extent at which systems in the Designed class adhere to statistical normality. Recall that for the purpose of this work, statistical normality is defined as the compounding effect of (a) the difference between actual and expected subgraph concentration and (b) the existence of statistical correlations between these patterns.

Focusing on component (a) of statistical normality, consider the QQ plots found at the diagonal of Fig.5 matrix plot, where the y-axis represents the expected subgraph concentration of the four most frequent subgraphs per sub-class, and x-axis corresponding to the actual subgraph concentration. The Construction Project sub-class exhibits limited dispersion across the diagonal, an indication of convergence between observed and expected value with the notable exception of m2 concentration. With respect to the Software sub-class, similar uniformity is observed, with the notable exception of m1 which is responsible for significant deviations. Conversely, the PCB sub-class 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 (see following paragraph). Such increased dispersion can have important practical implications. For example, 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, as actual subgraph concentrations tend to deviate from expected values (in the case of PCBs, this effect is especially pronounced), 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 seemingly equivalent systems.

Focusing on component (b) of statistical normality, the upper triangular of the matrix plot presents color-coded Spearman Correlation coefficients with respect to the relationship between each pair of subgraphs (with the lower triangular illustrating the actual relationships). All three subclasses exhibit a statistically significant $(p \leq 0.01)$ correlation between m1 and m2. 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 m2 concentration will fuel expectations of noting a reduction in the m3 concentration within the PCB sub-class, assuming all other variables remain constant. By inspection, one can also note that PCBs have the highest *average* $R^2$, though in absolute terms, it is relatively low, indicating 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.

In summary, all 3 sub-classes of the Designed class show increasingly low levels of adherence to statistical normality. Such behavior is typically described by deviations between actual and expected subgraph concentrations, along with weak correlations across possible subgraph combinations. As a result, Hypothesis 2 can be falsified.

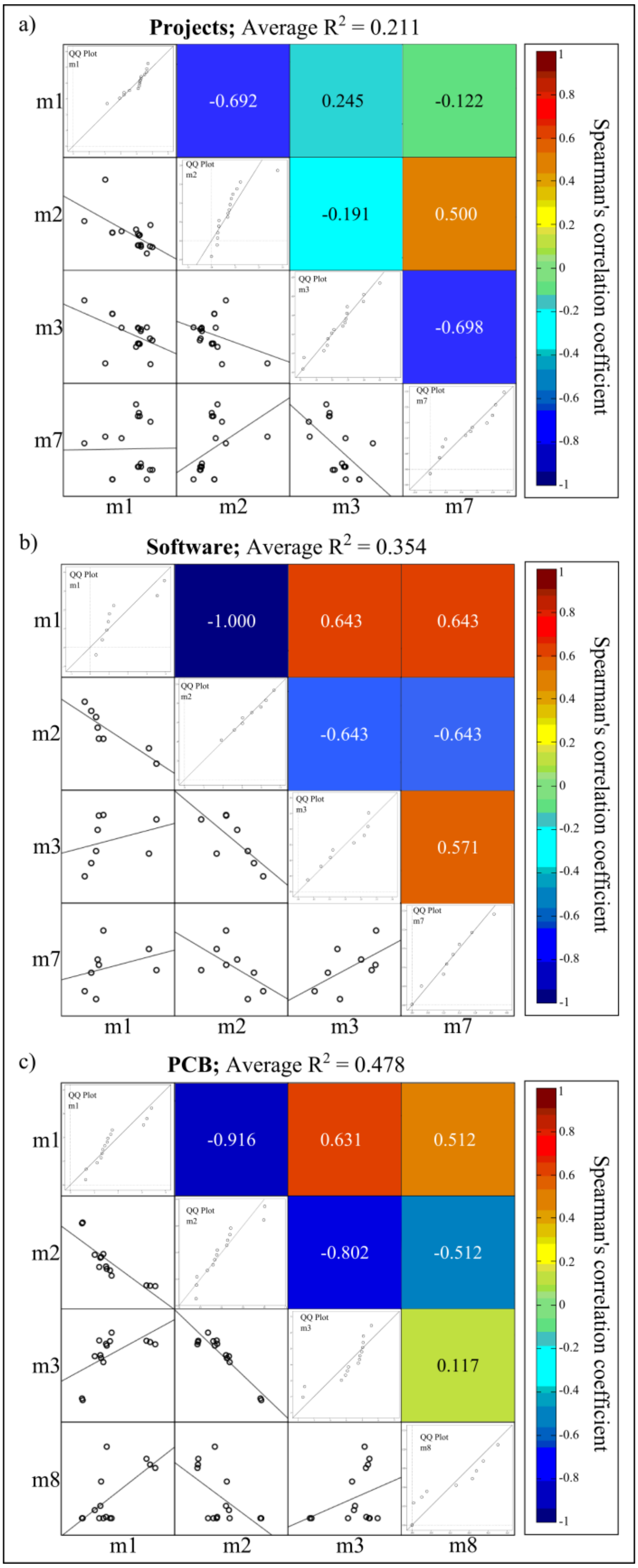

Fig.6. Matrix scatter plots for each of the three Designed sub-classes. The lower triangular section plots the relationship between all possible combinations between the four most

frequent subgraphs – their respective Spearman Coefficient is color-coded on the upper triangular part, ranging from -1 (perfect, negative, statistical dependence) to +1 (perfect, positive, statistical dependence). The diagonal section presents QQ plots where the expected subgraph concentration (y-axis) is plotted against the actual subgraph concentration (x-axis).

## V. Limitations and Future Work

This work has some limitations that provide opportunities for further work. Firstly this work has imposed a static view on the structure of the examined systems. Even through such view makes the analysis increasingly tractable, it is clearly a simplification. An increasingly realistic view is one where such structure is adaptive i.e. a change in the function of the system feeds back to its structure, forcing the structure to adapt [Gross and Blasius, 2008]. Future work could leverage the initiative of this work to explore the implications and underlying drivers of such adaptive behavior. As an example, consider the temporal data available for some of the Construction projects considered herein. Further work could focus on identifying structural difference between them and reasoning on the underlying process that led to these changes e.g. a change in requirements led to new tasks and dependencies emerging, subsequently affecting the structure of the underlying network.

A second limitation of this work is the absence of a generative model which can explain *why* these structural similarities and/or differences across these systems arise. This lack arises from the explicit focus of this work on mapping the structure of these systems – an important perquisite for developing such generative models. Building on the evidence presented herein, future studies may deploy machine-learning techniques to infer plausible generative process which account for the extent in which contextual limitations apply for a given system (i.e. limited resources, need for local/global optimization) as a possible driver for the different structural profiles noted between Designed and Evolved Systems.

## VI. Conclusions

This paper adopts a context agnostic view in order to compare the structural patterns of two general classes of systems: Designed and Evolved. Depending on the level of analysis, the emergence of patterns across seemingly different systems challenges the traditional view adopted within the engineering regime.

In terms of results, this work highlights pronounced patterns within both Designed and Evolved classes, both between and within these classes. Such patterns can be leveraged to potentially uncover the structural patterns that dictate the behavior of such systems e.g. Fig. 4 highlights two distinct behavior in terms of the capacity of each class to be controlled – a highly desirable property for any Designed system.

Examples of such work 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.

## VII. Acknowledgments

CE acknowledges the financial support by the EPSRC funded Industrial Doctorate Centre in Systems (Grant EP/G037353/1) and Systemic Consult Ltd. Data contribution by LOIS Builders Ltd is gratefully acknowledged. CE thanks the developers of MAVISTO for making it freely available, Dr. Babak Heydari for the invitation in contributing to the Journal and the two anonymous referees. CE, NA and AJ designed the research; CE developed the method, analyzed the results and wrote the manuscript; CE, NA and AJ reviewed the manuscript.

## VIII. References

1. W.B. Arthur, The nature of technology: What it is and how it evolves, Free Press, 2009.
2. W. Weaver, Science and complexity, American scientist 36 (1948), 536-544.
3. R. Pastor-Satorras, and A. Vespignani, Epidemic spreading in scale-free networks, Physical Review Letters 86 (2001), 3200.
4. A. Vespignani, Modelling dynamical processes in complex socio-technical systems, Nature Physics 8 (2012), 32-39.
5. J. Lorenz, S. Battiston, and F. Schweitzer, Systemic risk in a unifying framework for cascading processes on networks, The European Physical Journal B 71 (2009), 441-460.
6. Y.-Y. Liu, J.-J. Slotine, and A.-L. Barabási, Observability of complex systems, Proceedings of the National Academy of Sciences 110 (2013), 2460-2465.
7. J. Sterman, Business dynamics, Irwin-McGraw-Hill, 2000.
8. Y. Ma, K. He, and D. Du, A qualitative method for measuring the structural complexity of software systems based on complex networks, 2013 20th Asia-Pacific Software Engineering Conference (APSEC), 2005, pp. 257-263.
9. M.d. Magistris, M.d. Bernardo, and C. Petrarca, Experiments on synchronization in networks of nonlinear oscillators with dynamic links, Nonlinear Theory and Its Applications, IEICE 4 (2013), 462-472.
10. C. Ellinas, N. Allan, C. Durugbo, and A. Johansson, How robust is your project? From local failures to global catastrophes: A complex networks approach to project systemic risk, PLoS ONE 10 (2015), e0142469.
11. Y. Bar-Yam, When systems engineering fails-toward complex systems engineering, Systems, Man and Cybernetics, 2003. IEEE International Conference on, vol. 2, 2003, pp. 2021-2028.
12. ICCPM, "Global perspectives and the strategic agenda to 2025," *Complex Project Management*, International Centre for Complex Project Management, Australia, 2011.
13. A. Punter, "Supply chain failures: A study of the nature, causes and complexity of supply chain disruptions," airmic, London, 2013.

14. V. Venkatasubramanian, Systemic failures: Challenges and opportunities in risk management in complex systems, AIChE Journal 57 (2011), 2-9.

15. A. Vespignani, Complex networks: The fragility of interdependency, Nature 464 (2010), 984-985.

16. T.M. Williams, Modelling complex projects, John Wiley & Sons, 2002.

17. R.M. May, S.A. Levin, and G. Sugihara, Complex systems: Ecology for bankers, Nature 451 (2008), 893-895.

18. F. Schweitzer, G. Fagiolo, D. Sornette, F. Vega-Redondo, and D.R. White, Economic networks: What do we know and what do we need to know?, Advances in Complex Systems 12 (2009), 407-422.

19. A.G. Haldane, and R.M. May, Systemic risk in banking ecosystems, Nature 469 (2011), 351-355.

20. D. Helbing, Globally networked risks and how to respond, Nature 497 (2013), 51-59.

21. A.-L. Barabási, The network takeover, Nature Physics 8 (2011), 14.

22. M.A. Bellamy, and R.C. Basole, Network analysis of supply chain systems: A systematic review and future research, Systems Engineering 16 (2013), 235-249.

23. S.A. Sheard, and A. Mostashari, Principles of complex systems for systems engineering, Systems Engineering 12 (2009), 295-311.

24. PMI, "A guide to the project management body of knowledge: Pmbok® guide," Project Management Institute, 2008.

25. Y. Bar-Yam, Dynamics of complex systems, Addison-Wesley Reading, MA, 1997.

26. J.M. Ottino, Complex systems, AIChE Journal 49 (2003), 292-299.

27. ---, Engineering complex systems, Nature 427 (2004), 399-399.

28. C. Ellinas, N. Allan, and A. Johansson, Toward project complexity evaluation: A structural perspective, IEEE Systems Journal PP (2016), 1-12.

29. B. Heydari, and K. Dalili, Emergence of modularity in system of systems: Complex networks in heterogeneous environments, Systems Journal, IEEE 9 (2015), 223-231.

30. G.A. Polacek, D.A. Gianetto, K. Khashanah, and D. Verma, On principles and rules in complex adaptive systems: A financial system case study, Systems Engineering 15 (2012), 433-447.

31. S.A. Sheard, and A. Mostashari, Principles of complex systems for systems engineering, Syst. Eng. 12 (2009), 295-311.

32. V. Paxson, and S. Floyd, Wide area traffic: The failure of poisson modeling, IEEE/ACM Transactions on Networking (ToN) 3 (1995), 226-244.

33. T. Tyrakowski, and Z. Palka, A random graph model of mobile wireless networks, Electronic Notes in Discrete Mathematics 22 (2005), 311-314.

34. C. Ellinas, N. Allan, and N. Cantle, How resilient is your organisation? From local failures to systemic risk, ERM Symposium 2015, 2015.

35. B.B. Mandelbrot, and R.L. Hudson, The misbehavior of markets: A fractal view of financial turbulence, Basic books, 2007.

36. P.W. Anderson, More is different, Science 177 (1972), 393-396.

37. M. Geli-Mann, Complex adaptive systems, Complexity: Metaphors, models and reality (1994), 17-45.

38. A.-L. Barabási, The architecture of complexity, Control Systems, IEEE 27 (2007), 33-42.

39. M.E.J. Newman, Resource letter cs–1: Complex systems, American Journal of Physics 79 (2011), 800-810.

40. D.J. Watts, and S.H. Strogatz, Collective dynamics of ‘small-world’networks, Nature 393 (1998), 440-442.

41. B. Barzel, and A.-L. Barabási, Universality in network dynamics, Nature physics 9 (2013), 673-681.

42. A.-L. Barabási, and R. Albert, Emergence of scaling in random networks, Science 286 (1999), 509-512.

43. W. Willinger, R. Govindan, S. Jamin, V. Paxson, and S. Shenker, Scaling phenomena in the internet: Critically examining criticality, Proceedings of the National Academy of Sciences 99 (2002), 2573-2580.

44. S.P. Borgatti, A. Mehra, D.J. Brass, and G. Labianca, Network analysis in the social sciences, Science 323 (2009), 892-895.

45. L.-K. Chan, and M.-L. Wu, Quality function deployment: A literature review, European Journal of Operational Research 143 (2002), 463-497.

46. INCOSE, Incose systems engineering handbook: A guide for system life cycle processes and activities, John Wiley & Sons, 2015.

47. B.V. Koen, Definition of the engineering method, ERIC, 1985.

48. M.E. Newman, Networks: An introduction, Oxford University Press, 2009.

49. P. Godfrey, "Architecting complex systems in new domains and problems: Making sense of complexity and managing its unintended consequences," Complex systems design & management, Springer, 2013, pp. 41-51.

50. M. Mitchell, Complex systems: Network thinking, Artificial Intelligence 170 (2006), 1194-1212.

51. R.C. Basole, W.B. Rouse, L.F. McGinnis, D.A. Bodner, and W.C. Kessler, Models of complex enterprise networks, Journal of Enterprise Transformation 1 (2011), 208-230.

52. D. Lazer, A.S. Pentland, L. Adamic, S. Aral, A.L. Barabasi, D. Brewer, N. Christakis, N. Contractor, J. Fowler, and M. Gutmann, Life in the network: The coming age of computational social science, Science (New York, NY) 323 (2009), 721.

53. A.-L. Barabási, Scale-free networks: A decade and beyond, Science 325 (2009), 412-413.

54. M.E.J. Newman, The structure and function of complex networks, SIAM review 45 (2003), 167-256.

55. R. Albert, and A.-L. Barabási, Statistical mechanics of complex networks, Reviews of modern physics 74 (2002), 47.

56. S. Boccaletti, V. Latora, Y. Moreno, M. Chavez, and D.-U. Hwang, Complex networks: Structure and dynamics, Physics reports 424 (2006), 175-308.

57. S.N. Dorogovtsev, and J.F. Mendes, Evolution of networks, Advances in physics 51 (2002), 1079-1187.

58. P. Erdos, and A. Rényi, On the evolution of random graphs, Publ. Math. Inst. Hungar. Acad. Sci 5 (1960), 17-61.

59. S.H. Strogatz, Exploring complex networks, Nature 410 (2001), 268-276.

60. D.J. Watts, The" new" science of networks, Annual review of sociology (2004), 243-270.

61. M. Karsai, M. Kivelä, R.K. Pan, K. Kaski, J. Kertész, A.-L. Barabási, and J. Saramäki, Small but slow world: How network topology and burstiness slow down spreading, Physical Review E 83 (2011), 025102.

62. D. Centola, V.M. Eguíluz, and M.W. Macy, Cascade dynamics of complex propagation, Physica A: Statistical Mechanics and its Applications 374 (2007), 449-456.

63. V. Latora, and M. Marchiori, Efficient behavior of small-world networks, Physical Review Letters 87 (2001), 198701.

64. X.F. Wang, and G. Chen, Synchronization in small-world dynamical networks, International Journal of Bifurcation and Chaos 12 (2002), 187-192.

65. R. Milo, S. Shen-Orr, S. Itzkovitz, N. Kashtan, D. Chklovskii, and U. Alon, Network motifs: Simple building blocks of complex networks, Science 298 (2002), 824-827.

66. R. Milo, S. Itzkovitz, N. Kashtan, R. Levitt, S. Shen-Orr, I. Ayzenshtat, M. Sheffer, and U. Alon, Superfamilies of evolved and designed networks, Science 303 (2004), 1538-1542.

67. U. Alon, Network motifs: Theory and experimental approaches, Nature Reviews Genetics 8 (2007), 450-461.

68. P.J. Ingram, M.P. Stumpf, and J. Stark, Network motifs: Structure does not determine function, BMC genomics 7 (2006), 108.

69. J.F. Knabe, C.L. Nehaniv, and M.J. Schilstra, Do motifs reflect evolved function?—no convergent evolution of genetic regulatory network subgraph topologies, Biosystems 94 (2008), 68-74.

70. O. Sporns, and R. Kötter, Motifs in brain networks, PLoS biology 2 (2004), e369.

71. S. Lämmer, and D. Helbing, Self-stabilizing decentralized signal control of realistic, saturated network traffic, 2010.

72. F. Brglez, D. Bryan, and K. Kozminski, Combinational profiles of sequential benchmark circuits, Circuits and Systems, 1989., IEEE International Symposium on, 1989, pp. 1929-1934.

73. F. Brglez, and R. Drechsler, Design of experiments in cad: Context and new data sets for iscas'99, Circuits and Systems, 1999. ISCAS'99. Proceedings of the 1999 IEEE International Symposium on, vol. 6, 1999, pp. 424-427.

74. C.R. Myers, Software systems as complex networks: Structure, function, and evolvability of software collaboration graphs, Physical Review E 68 (2003), 046116.

75. L. Ying, and D.-w. Ding, Topology structure and centrality in a java source code, Granular Computing (GrC), 2012 IEEE International Conference on, 2012, pp. 787-789.

76. D. Auber, D. Archambault, R. Bourqui, A. Lambert, M. Mathiaut, P. Mary, M. Delest, J. Dubois, and G. Mélançon, The tulip 3 framework: A scalable software library for information visualization applications based on relational data, (2012).

77. R. Guimera, L. Danon, A. Diaz-Guilera, F. Giralt, and A. Arenas, The real communication network behind the formal chart: Community structure in organizations, Journal of Economic Behavior & Organization 61 (2006), 653-667.

78. V. Batagelj, and A. Mrvar, Pajek datasets, Web page http://vlado.fmf.uni-lj.si/pub/networks/data (2006).

79. L.A. Adamic, and N. Glance, The political blogosphere and the 2004 us election: Divided they blog, Proceedings of the 3rd international workshop on Link discovery, 2005, pp. 36-43.

80. W.W. Zachary, An information flow model for conflict and fission in small groups, Journal of anthropological research (1977), 452-473.

81. H. Jeong, S.P. Mason, A.-L. Barabási, and Z.N. Oltvai, Lethality and centrality in protein networks, Nature 411 (2001), 41-42.

82. J. Bascompte, 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 102 (2005), 5443-5447.

83. M. Vanhoucke, An overview of recent research results and future research avenues using simulation studies in project management, ISRN Computational Mathematics 2013 (2013).

84. A. Sharon, O.L. de Weck, and D. Dori, Project management vs. Systems engineering management: A practitioners' view on integrating the project and product domains, Systems Engineering 14 (2011), 427-440.

85. L.-A. Vidal, and F. Marle, Understanding project complexity: Implications on project management, Kybernetes 37 (2008), 1094-1110.

86. A.J. Shenhar, One size does not fit all projects: Exploring classical contingency domains, Management Science 47 (2001), 394-414.

87. D. Baccarini, The concept of project complexity—a review, International Journal of Project Management 14 (1996), 201-204.

88. V. Valls, and P. Lino, Criticality analysis in activity-on-node networks with minimal time lags, Annals of operations research 102 (2001), 17-37.

89. S. Wasserman, Social network analysis: Methods and applications, Cambridge university press, 1994.

90. F. Schreiber, 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), 91-102.

91. F. Schreiber, and H. Schwöbbermeyer, Mavisto: A tool for the exploration of network motifs, Bioinformatics 21 (2005), 3572-3574.

92. S. Valverde, and R.V. Solé, Network motifs in computational graphs: A case study in software architecture, Physical Review E 72 (2005), 026107.

93. Y.-Y. Liu, J.-J. Slotine, and A.-L. Barabási, Controllability of complex networks, Nature 473 (2011), 167-173.

94. J.M. Carlson, and J. Doyle, Complexity and robustness, Proceedings of the National Academy of Sciences 99 (2002), 2538-2545.

95. T. Gross, and B. Blasius, Adaptive coevolutionary networks: A review, Journal of the Royal Society Interface 5 (2008), 259-271.

## IX. Appendix

As previously defined, matrix $\mathbf{A}$ (adjacency matrix) is a $N \times N$ matrix that captures the underlying topology of a network. Furthermore, $\mathbf{B}$ is an $N \times M$ (where $M \leq N$) matrix that indicates which nodes are controlled by an outside signal. Using this notation, [Liu, Slotine and Barabási [2011]] have proposed that a network is controllable if its controllability matrix

$$\mathbf{C} = [\mathbf{B}, \mathbf{AB}, \mathbf{A}^2\mathbf{B}, \ldots, \mathbf{A}^{N-1}\mathbf{B}] \qquad (1A)$$

has full rank i.e. rank $\mathbf{C} = N$.

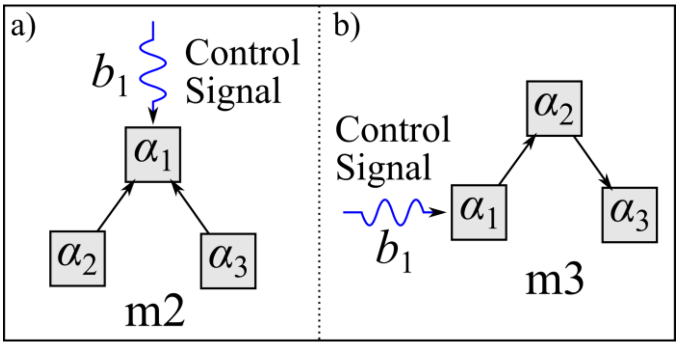

Fig. 1A: Figure captures an attempt to control (a) m1 and (b) m2 by enforcing an external signal ($b_1$) on the lead node, $a_1$.

With reference to Figure 1A(a), the adjacency matrix $\mathbf{A}$ of m2 is given by:

$$\mathbf{A} = \begin{bmatrix} 0 & 1 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

with the controllability matrix $\mathbf{C}$ defined as:

$$\mathbf{C} = [\mathbf{B}, \mathbf{A} \cdot \mathbf{B}, \mathbf{A}^2\mathbf{B}] = \left[ \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 & 1 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \cdot \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 & 1 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}^2 \cdot \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} \right] = \begin{bmatrix} 0 & 1 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

Since rank $\mathbf{C} = 1 < N$, m2 is uncontrollable. Note that this result holds even if the control input is exerted on another node i.e. $\mathbf{B} = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}$ or $\mathbf{B} = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}$.

In the case of m3, its structure is defined as:

$$\mathbf{A} = \begin{bmatrix} 0 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}$$

with the controllability matrix $\mathbf{C}$ defined as:

$$\mathbf{C} = [\mathbf{B}, \mathbf{AB}, \mathbf{A}^2\mathbf{B}] = \left[ \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix} \cdot \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}^2 \cdot \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} \right] = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

with rank $\mathbf{C} = 3 = N$, m3 can be considered to be structurally controllable.

Finally, note that in this case, both m2 and m3 are treated as unweighted – nonetheless these results also hold for their weighted instances. Further details can be found in the Supplementary Information, Section III of [Liu, Slotine and Barabási, 2011].