PAPER REF # 8242
Proceedings: Eighth International Space Syntax Symposium
Edited by M. Greene, J. Reyes and A. Castro. Santiago de Chile: PUC, 2012.

ORDER, STRUCTURE AND DISORDER IN SPACE SYNTAX AND LINKOGRAPHY: intelligibility, entropy, and complexity measures

AUTHOR: **Tamer EL-KHOULY**
Bartlett School of Graduate Studies, UCL Faculty of the Built Environment, University College London
UCL, United Kingdom
e-mail: t.el-khouly@ucl.ac.uk

Alan PENN
Bartlett School of Graduate Studies, UCL Faculty of the Built Environment, University College London
UCL, United Kingdom
e-mail: a.penn@ucl.ac.uk

KEYWORDS: *Linkography, Order, Disorder, Strings of information, Intelligibility, Complexity, Entropy*

THEME: Methodological Development and Modeling

Abstract

There has been great interest in the use of linkographies to describe the events that take place in design processes with the aim of understanding when creativity takes place and the conditions under which creative moments emerge in the design. Linkography is a directed graph network and because of this it gives resemblance to the types of large complex graphs that are used in the space syntax community to describe urban systems. In this paper, we investigate the applications of certain measures that come from space syntax analyses of urban graphs to look at linkography systems. One hypothesis is that complexity is created in different scales in the graph system from the local sub-graph to the whole system. The method of analysis illustrates the underlying state of any system. Integration, complexity and entropy values are measured at each individual node in the system to arrive at a better understanding on the rules that frame the relationships between the parts and the whole.

INTRODUCTION

A linkography is a representation of the series of events that can be observed to occur and can be used to help analyse processes of creativity during a design session. The difference between linkography and spatial system is that linkography has a time factor. A linkograph is constructed from nodes that represent each segment in the design process (according to time) and is based on parsing the dependency relationships between those nodes.

Because it is a representation that traces the associations of every single utterance, the design process can be looked at in terms of a linkographic pattern that displays the structure of the design reasoning. The venues of dense interrelations (clusters of the design utterances) are overtly highlighted on the graph and can be further interpreted through the emerging artefacts along the process.

The linkography system is hypothesised to deliver a variation of complexity degrees on different occasions. The aim is to uncover the significant events that might be associated with creative insights and inspect the artefacts that are formulated at such events. Linkography and urban systems deal with multi-level complexities, the overall goal of the proposed analytical method is to reveal the relationship between the parts (sub-networks) that constitute the system and the whole.

The relationship between the sub-systems or the partial assemblies is inspected looked at from two perspectives, information theory and entropy theory, to see whether a conflict occurs between uncoordinated sub-orders despite being orderly structured (Arnheim, 1971; Laing, 1965) or whether an order system underlies an entire disorder state (Planck, 1969) – an entity that is dependent on a random dispersion of limited sub-orders (Arnheim, 1971; Kuntz, 1968).

A computational model is proposed that covers the dependency relationships occurring between nodes, all of which appear to have a sophisticated group of relations. The algorithm used is inspired by the T-code string measure developed by Titchener (1998a; 1998b; 1998c; 2004).

1. A POINT OF DEPARTURE

A gridiron urban system is perceived as a highly organised structure if it delivers different chances to navigate from one place to another. It is highly intelligible in this circumstance, but to some extent it can become confusing. In a very symmetrical and identical system, the explorer has equal chances to move from one point in the system to another and might get lost. Since intelligibility is the correlation between connectivity and integration, hence the same correlation value is constituted for any element in this particular system.

In reality, no system is perfectly set up as a 100% identical gridiron. Every city has some sense of differentiation that adds to the structure and provides the capacity to grasp the relation between the "whole" and the "parts". The example of two forests, natural-spontaneous and farmed-grid, reflects two different states. In the first case, trees are not aligned and the distribution is chaotic, while in the second, trees are strictly planted along straight lines and the arrangement is similar everywhere in the network. In both cases, systems disorientate the explorer from a proper navigation. Yet the highly ordered forest and the disordered one are both considered extreme examples in terms of intelligibility.

What deserves attention is how we construct a system; a city for instance. Hanson (1989) has pointed that order might be misleading about its function and that it could be a manifestation of another underlying state. Hence, the importance of distinguishing this kind of relationship is crucial to reveal the real state at each stage of a multi-level complex system. What we mean is that something might occur on the system midway between total chaos and total order, a certain point where it starts to behave differently from the preceding state(s).

The demonstration of the gridiron order, despite being a singularity in intelligibility terms, is as unintelligible as the total chaos state. Both systems deliver lack of intelligibility for the thinking subject. Order, in this particular case, is just the same as complete disorder in delivering a lack of intelligibility. However, if we impose a differentiation on the gridiron by adding some diagonals and routes, the whole structure has not drastically changed but its intelligibility moves from one state to another (the system becomes more intelligible otherwise).

In fact, working with systems that have multi-level complexity on different scales is common in urban and linkography systems. One view is that there is a clear order and that the structure of the system can be easily grasped and understood. The other view is that there is no rule in the complex world and that it is actually just random. The paradox is that if it is truly random is there a simple way to describe it? Can a complex world be reduced to a single value?

This paper proposes a hypothesis that in multi-level complex systems high orderliness tends to become less complex overall, and therefore a highly linked node delivers few choices and probabilities. The alternative to inspecting the system is therefore to measure the probability for each node and complexity at each level (at every sub-network) included within the system. In doing so, we propose the adoption of strings of information to code probabilities at each point and compute the information content from it. The practical aims of using this method are twofold.

First, since all the inspected sub-networks have the same sub-graph size effect, the measures of strings at each point in the system are already relativised and eligible for comparison. This is because the information is extracted for all the possible relations that could be made from any point in the system to the others (the sub-graph size always equals n-1).

Second, integration values are also relativised to the sub-graph size. Thus, integration, complexity, rate and content of information are relativised parameters that we look at in order to specify the relation between the parts constituting the whole.

2. ENTROPY AND INFORMATION

Space syntax and design process are multi-scaled complex contexts. The information content at different scales reflects the complexity at each level in the system. In the proposed method, the system can be read in two ways. The first looks at the probability of choice at any “item”, “point”, or “node”, while the second looks at the rate of information measured for a “sequence” of items.

The methods correspond to entropy theory and information theory respectively. But while entropy is concerned with “sets” of individual items, information is concerned with the individual “sequence” of those items. The entropy theory asserts that a “set” should be treated as a “microstate”; the microstates constitute the complexions of the overall process.[1] At this point, the main object of inquiry in information

[1] Arnheim (1971) described the microstate in the principle of entropy theory as: “the particular character of any microstate does not matter; its structural uniqueness, orderliness or disorderliness does not count. What does matter is the *totality* of these innumerable complexions adding up to a global macrostate of the whole process. It is *not* concerned with the probability of succession in a series of items but with the *overall distribution* of kinds of items in a given arrangement.”

theory is to investigate the probability of occurrence by establishing the number of possible sequences. The "sequence" of items is not covered in entropy theory but is necessary for information theory. Table 1 illustrates the differences that distinguish the two perspectives.

	Entropy theory	Information theory
Structure	▪ Items constitute the main characteristics of the structure	▪ Structure means nothing is better than those certain "sequences" of items that can be expected to occur
Central points	▪ Concerned with "sets" of individual items ▪ Is about the "overall distribution" of kinds of items in a given arrangement ▪ The more remote the arrangement of sets is from a random distribution, the lower its entropy, and the higher its order representation	▪ Focused on the individual "sequence" of items ▪ Is about "sequences" and "arrangements" of item. ▪ The less predictable the sequence, the more information the sequence will yield, and the more remote its representation from order
Example	▪ A randomised distribution will be called by the entropy theorist "highly probable" and therefore of low order because innumerable distributions of this kind can occur	▪ A highly randomised sequence will be said to carry "much information" by the information theorist because information in this sense is concerned with the probability of this particular sequence
Application	▪ For example, Kan and Gero's (2005a; 2007; 2008) estimation method to acquire entropy from linkography ▪ For example, Turner's (2007) adoption of Shannon's formula to estimate entropy for urban systems with *Depthmap*	▪ For example, Titchener et al.'s (2005) computation of strings of information ▪ For example, Brettel's (2006) adoption of Titchener's (2004) t-code measures to estimate entropy for navigation routes

Table 1: The differences that distinguish entropy theory and information theory

An observer would find that the most highly ordered system provides maximum information content and thus is opposite to probabilistic entropy since the prediction is very high. If total disorder provides maximum information as well, then maximum order is conveyed by maximum disorder (Arnheim, 1971). However the distinction can be made through a parameter that measures the underlying system of any order. Since information is a crude measure that confirms a clear increase in regularity overall, extreme regularity and apparent similarity are likely to deliver a very low probability value.

Entropy grows with the probability of a state of affairs while information does the opposite and increases with the improbability. The less likely an event is to happen, the more information its occurrence represents. The least predictable sequence of events will carry the maximum information. Hence, this paper focuses on how entropy could be estimated for multi-level systems in a way that views the relationship between the nature of complexion between the partial assemblies that are made at each point and the whole. The proposed method therefore adopts entropy and complexity as independent measures to assess complex systems such as linkography. However, it should be noted that the structure state of any system needs a variation of characteristics in order to construct an intelligible system.[2] The next section reviews methods to estimate entropy and introduces the computational method of strings of information.

[2] Referring back to the example of the gridiron system, all elements deliver the same correlation value between connectivity and integration, however any imposed differentiation on the gridiron cause variations on intelligibility, and then system changes from a state to another.

3. ENTROPY OF SPATIAL SYSTEMS AND LINKOGRAPHY

The estimation of entropy for spatial systems is based on the frequency distribution of the point depths (Turner, 2007). The point depth entropy of a location, s_i, is expressed by utilising Shannon's formula of uncertainty as shown in the equation:

Point Depth Entropy for spatial system →

$$s_i = \sum_{d=1}^{d_{max}} - P_d \log_2 P_d \quad \text{… … …} \quad \textbf{(I)}$$

Where d_{max} is the maximum depth from vertex v_i and P_d is the frequency of point depth d from the vertex

Estimating point depth entropy in this way shows how orderly a spatial system is structured from a certain location. The method is a functional equation based on "mean depth". *In Depthmap*, the information from a point is calculated with respect to the expected frequency of locations at each depth. Turner (2007) explained that the "expected" frequency is based on the probability of events occurring depending on a single variable, the "mean depth" of the j-graph. The benefit of calculating *entropy* or *information* from a "point" in space syntax pertains to how easy it is to traverse to a certain depth within the system. Low disorder is easy; high disorder is hard.

In linkography, with referral to (Gero et al.; 2011, Kan and Gero, 2005b; 2005c; 2007; 2008; 2009a; 2011b; and Kan et al.; 2006; 2007), Shannon's theory of information (1949) is adopted to inspect the occurrence of dependency relationships between utterances.[3] This gives two possible choices to code the system: "linked" and "unlinked" (or "on" and "off"). The formula used is:

Shannon Entropy for Linkography →

$$H = -(p_{linked} . \log_2 p_{linked}) + (p_{unlinked} . \log_2 p_{unlinked}) \quad \text{… … …} \quad \textbf{(II)}$$

Kan and Gero's method looks at the overall distribution of "sets" (items of relations) regardless of the "sequence" of occurrence of elements constituting the linkography according to time. The example in Figure 1 emphasises that the differences between two linkographic patterns are not considered in the estimation process of entropy. This is owing to the summation step – processed over the whole network – for each of the two probabilities, "linked" and "unlinked", regardless of the position of items in the system that should precede the estimation.[4]

[3] Goldschmidt (1992) defined a design "move" or "step" in the following terms: "a move is an act of reasoning that presents a coherent proposition pertaining to an entity that is being designed". Goldschmidt (1995) also stated: "a step, an act, or an operation, which transforms the design situation relative to the state in which it was prior to that move". See also: Goldschmidt (1990).

[4] Remember that linkography is a directed graph according to time factor.

Example 1

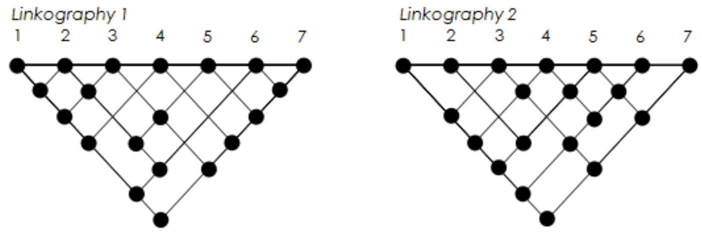


Figure 1: Two linkography patterns that are different in the arrangement of dependency relationships but have the same Shannon entropic value

Processing Shannon's Entropy on linkography:

$$\mathbf{H = -(p_{linked} \cdot \log_2 p_{linked}) + (p_{unlinked} \cdot \log_2 p_{unlinked})}$$

The total number of possible relationships = $n[(n-1)/2] = 7(6/2) = 21$

Where n is the total size of the linkography (the number of nodes).

The total number of "linked" relations in both graphs is = 13 → □ 61.9%

The total number of "unlinked" relations in both graphs is = 8 → □ 38.1%

$H = -[(13/21) \times (\log_2(13/21))] + [(8/21) \times (\log_2(8/21)]$

$H = -[(0.62) \times (-0.69)] + [0.38] \times (-1.39)]$

$H = -0.4 \times 0.5 = 0.2$ bit/bits

Both graphs have the same entropy value despite the clear difference of arrangements in each system. This is because the equation is based on summing the values of each probability without considering the position of each in the existing pattern. The next section provides a synopsis on intelligibility in space syntax. It illustrates a brief from a previous study (Brettel 2006) that combined string measures with integration values on spatial networks with distinctive configurations, investigating the connectivity between nodes through *navigation* in various samples.

4. THE COMPUTATION OF STRINGS OF INFORMATION

An inclination towards the hypothesis is delivered throughout Brettel's (2006) study, which investigated how "order", "structure", and "disorder" of street layouts are perceived when navigating through an urban environment. She stated that "an ordered environment tends to be more intelligible when broken up by an irregularity occasionally." In our study, we ask: *Under which circumstances does the system change from one state to another?* But more specifically on Brettel, we ask: *Are highly intelligible spatial systems predictable to navigate through?* and *Does simple traverse through urban fabric deliver less complex structure*?[5]

The string of information measures to deal with "event" structure was introduced in Brettel's study[6] in order to compute barcodes of event sequences extracted from navigation routes, in addition to syntactical analysis. The string measures were expected to relate to the perceived order along a route. The entropy of

[5] In other words, if the mechanism of access from one point to another is simple, does the synthesis form of its route deliver low complexity?

[6] An "event" is defined as a segment of time at a given location that is perceived by an observer to have a beginning and an end (Tversky and Zacks, 2001).

each route's string was interpreted as the probability of the uncertainty that a route provides for the traveller, and was expected to relate to the perceived structure along routes.[7] When a route has very few turns, the probability of choices is too low (for example, gridiron patterns such as New York and San Francisco). However, entropy delivers high values (relatively) with complex patterns when the route consists of some turns and deviations within it (for example, composite fabrics such as London and Rome). Moreover, the isovist fields owed the differentiation of visual catchment areas between the analysed cities not only according to the "delineation" in the route but also because of picking up structurally different catchment areas, especially in the irregular patterns.

According to Brettel's analyses, the computation process of strings could deliver meaningful correlations for the perceived route. Nevertheless, the assumption that orderliness is likely to be more related to complexity measure and structure to entropy could not be proven in her study, possibly due to limits of the survey setup.[8]

So are we measuring *"Intelligibility" versus "complexity" or "integration" versus "complexity"*? The central point of attention is to realise that intelligibility is a system property; the correlation between connectivity (C) and integration (I). Complexity (C_T) of graphs is also a system property that reflects how many steps are required to construct a string of information for the system (or sub-system). Consequently, values of the two parameters can be compared and correlated together. The sub-graph at each node is also a sub-system and the same measures can be used to inspect the characteristic within the whole.

4.1 The T-code Measure

The application of the "T-code" string computation method is based on the deterministic information theory that was developed by Titchener (1998a; 1998b; 1998c; 2004). In this method, an algorithmic process is applied to sets of information to compute the string measures, denoted as "T-complexity" and "T-entropy" (see also: Titchener, 2004; Titchener et al., 2005; Speidel et al., 2006; Speidel, 2008). The string signifies various types of information encoded into symbols.

If the string comprises a repeating sequence of one symbol only (one attribute), then entropy declines to zero value and the complexity structure of the string get lower, e.g. OOOOOOOOOOOOO, but if a string is composed of two or more symbols then the probability of appearance gets higher, e.g. LROoRRoRLOoLLLORLOooOoR. This means the complexity of string increases according to the size of the symbols and the composition.

The size of string is a crucial factor since longer strings give more accurate measurements than short ones. The complexity of string depends on the number of production steps that are required to construct this string (Titchener, 2004). Example 2 gives a tape of 100 digital codes that is spontaneously composed by typically duplicating one symbol code to another one, by processing the T-code measure:

[7] The probability of choices that could be made at decision points for directional turns. Accordingly, entropy describes how much information is there in a "signal" or "event".

[8] This result may be limited owing to the small size of samples and short strings.

Example 2:

1. In the case of repeating one symbol for 100 times:

0000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000

Length: 100 characters

T-COMPLEXITY	**T-INFORMATION**	**T-ENTROPY**
6.64 (taugs)	10.4 (nats)	0.104 (nats/char)
6.64 (taugs)	15.0 (bits)	0.150 (bits/char)

2. In the case of coding a string of two symbols (0 and 1) for 100 long:

1000011100011000000001100000001100000000000011111000001000001100000000011000000000100000000010000010

Length (chars): 100

T-COMPLEXITY	**T-INFORMATION**	**T-ENTROPY**
16.91 | 16.32 (taugs)	43.30 | 41.10 (nats)	0.43 | 0.411 (nats/char)
16.91 | 16.32 (taugs)	62.46 | 59.29 (bits)	0.62 | 0.592 (bits/char)

There are a variety of means to illustrate a network; see Figures 2 and 3 for some examples. The scope of this paper is not concerned with multiple representations to illustrate the system, but rather intends to understand the constituting force that attunes the components of it (such as what is beyond the links and relations between the nodes).

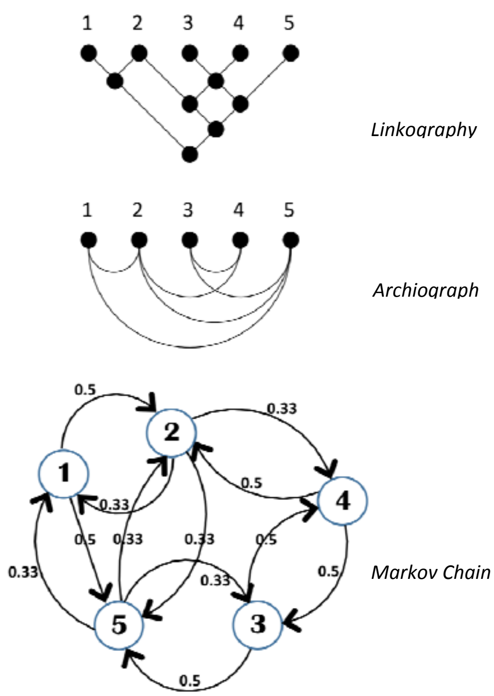


Figure 2: A variety of means to illustrate a system network: *Linkography, Archiography, and Markov Chain.*

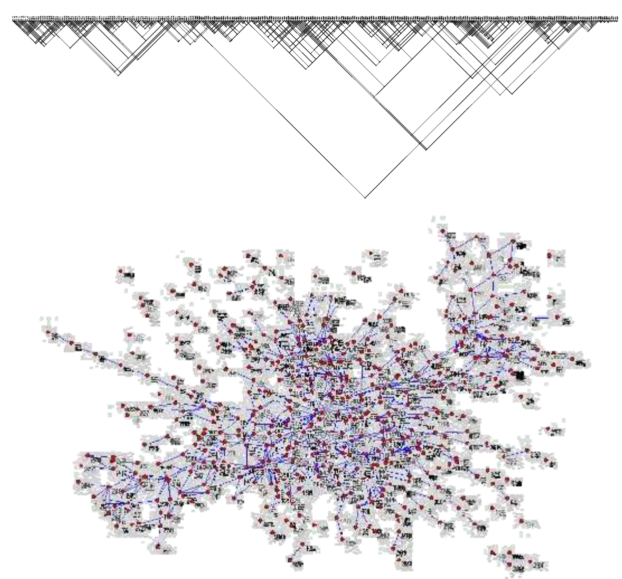


Figure 3: Two different representations for the same large system: a design protocol for one-hour design session consisting of 453 nodes.
Above is the linkography pattern and below is displayed upon the "connectivity strength" per nodes.

Before embarking on an analysis of the distribution of integration in each of the individual nodes in the system, we begin with a number of common features of the set of linkographies, which give some idea of the nature of the processes envisaged. After a preliminary study on some samples of linkography, the concluded points are twofold:

First, since the total number of links in any system of size n is (n-1), then the size of any node's possible relations equals (n-1) as well. This means that at any node, the sheer number of links in the sub-graph created from this node to the others in the system has the same size effect with every node. Accordingly, all the measures are relativised at every level in the system before embarking on comparisons. A second feature that differentiates between systems is the varied distribution of links. This should be considered in the estimation process of strings of information to include the sequences of sets in our interpretation rather than viewing the system at the node level only.

The linkography contains a structural hierarchy: the first level starts with the "nodes" that aggregate to form the "network" or "sub-system". Nodes and networks together construct the linkography pattern. It might happen in some cases that networks (sub-graphs) do not intersect together because the train of thoughts in this design venue is disconnected and the chunks of ideas are unrelated. However, in most cases networks intersect in one or more nodes. This means the design thoughts are structurally interrelated and built up. In this sense, linkography has different patterns and configurations: highly ordered/fully-saturated, structured, or disordered. There are some other configurations beyond these intrinsic types such as the mechanistic "saw-tooth" pattern that reflects a highly ordered and systematic (repetitive) process, or the "fully sparse" disconnected one that resembles a totally "unrelated" discourse, and so forth.

4.2 Processing the System as Multiple Sub-graphs

Any system can be transferred into strings of information by coding the dependency relations between nodes. In the case of linkography, a binary digit format is proposed: "1" for linked relationships and "0" for unlinked. The position of each symbol in the string refers to its sequence in the pattern according to time and thus the distribution of the dependency relationships is included while constructing the string. This section suggests a method to study the sub-graph at any node and extract the strings from it in order to compute the T-code measure on the linkography patterns.

Method: Computing the T-code measure on the sub-graph for each node

The estimation process is based on the concatenation of "back" and "fore" links together for each node in the system (see Figure 5). Despite the sub-graphs ("concatenation" of links per node) having equal sizes at every node, T-complexity and T-entropy values fluctuate along the linkography. An example of how to extract a string of information for a certain node's sub-graph is shown in Figure 6.

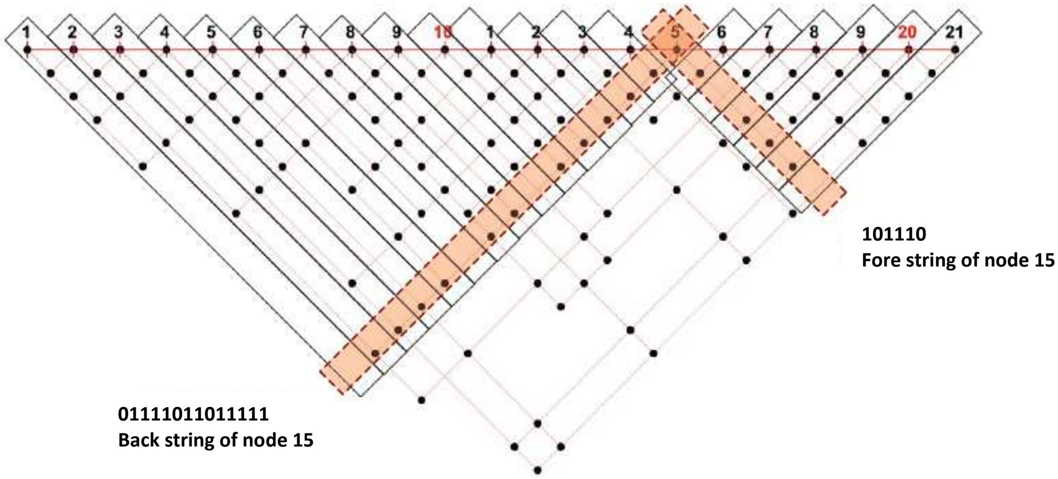


Figure 5: Processing dynamic T-complexity on complete linkography by concatenating the back and fore links together per node

Example 3:

According to Figure 5, the sub-graph created at node no.15 consists of the following relations:

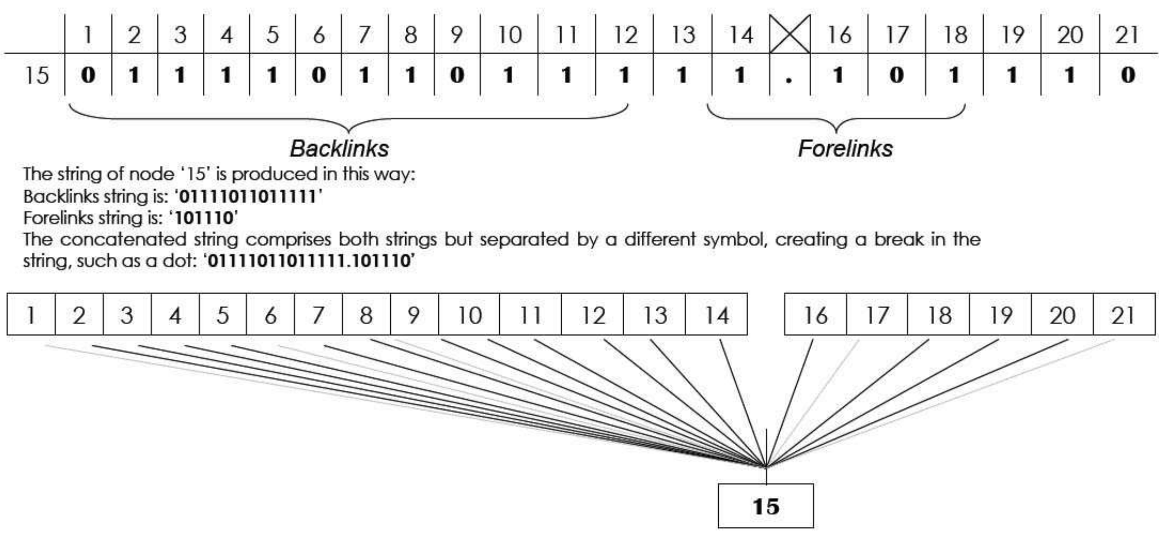


Figure 6: Example of extracting the string of information per sub-graph

4.3 Intelligibility, Complexity and Entropy

Intelligibility and *complexity* are properties of system. For a graph that consists of 100 nodes, each node will have two values: (1) intelligibility, which is the correlation between two values: connectivity and integration; and (2) complexity, which is measured for the "sub-graph" of relations at this particular node. Both measures have "size" effects. For intelligibility, where a system size "n" (n<50) this means intelligibility will tend to be high (for example, a small village with 50 links, paths or axes gives the range of values 0<0.5<1.0).

The system is “intelligible” if the correlation value is more than 0.5, and "unintelligible" if the value is less than 0.5).

String measures such as complexity, information content and entropy also have “size” effects. For a string of size “n”, (n<20), values are “inaccurate”. Accuracy for information and entropy is limited for short strings due to the approximation of bound by the logarithmic integral function (see Titchener, 2004). For a string of (n>20), the T-complexity we are looking at is a “sub-graph” of the whole linkography, namely those directly connected to node we are estimating.

Subsequently, the T-complexity and T-entropy measures are comparable to the integration value at that same node. Hence, the highly connected nodes at any system could be correlated to the string measures at the same node in order to investigate the proposed hypothesis. According to Brettel (2006), that intelligibility is signified throughout orderly systems.

As a rule of a thumb, the shortest line between two points is a straight line that has a first order synthesis form. A *piazza* is highly accessible from all its surrounding points (areas), the proposed path of navigation is clear and easier to travel, and thus the expectedness is high and the complexity is low. A *cul de sac* has a very low integration value in the system and not many options exist to approach it – only one access point. That makes it very complex to reach.

Giving an example of a particular spatial structure, Figure 7 illustrates two hypothetical network systems that are connected via only one node (resembling a bridge between two riverbanks), the real relative asymmetry value of this single node equals zero. Since integration and real relative asymmetry (RRA) are inversely correlated, this means that the most integrated point in the system is the highly linked node. Other nodes in each side are equivalent in integration and RRA values (see Figure 7).

This network will still be represented in several ways. It is obvious to infer that both network sides are highly ordered. However the string of information for each node in the system contains repeated symbols that indicate only one possible option (symbol) of interconnectivity inside each side. This will significantly affect the computed barcode measures for each node in the system. For example, taking node ‘9’:

String processed: ‘11111111111111111111’

Length (chars): 20

T-COMPLEXITY	**T-INFORMATION**	**T-ENTROPY**
4.32 (taugs)	5.4 (nats)	0.274 (nats/char)
4.32 (taugs)	7.9 (bits)	0.396 (bits/char)

(N.B. Accuracy for information and entropy is limited for short strings, due to approximation of bound by the logarithmic integral function, li().

Example 4:

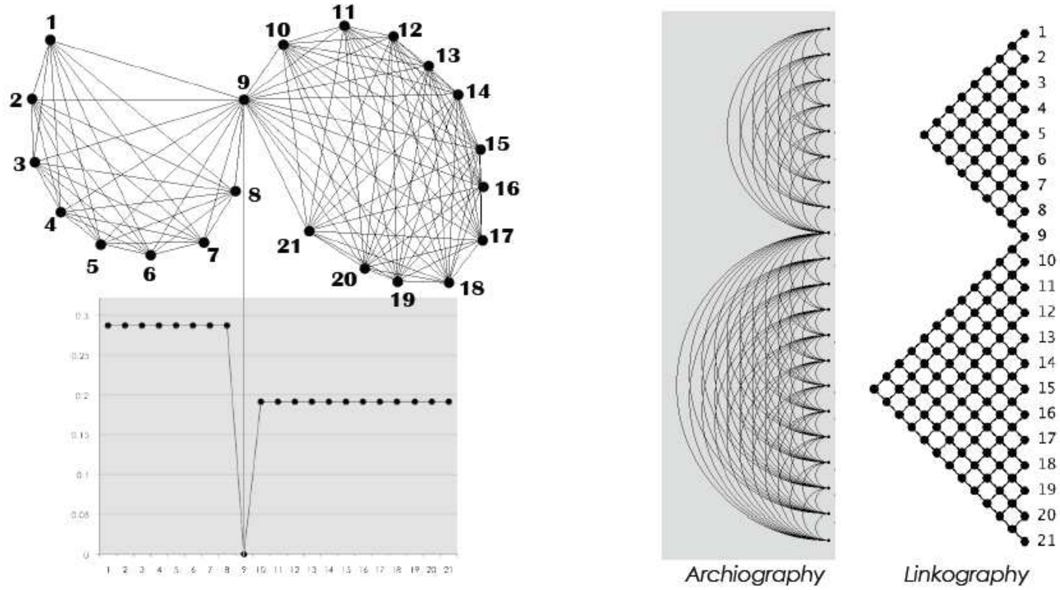
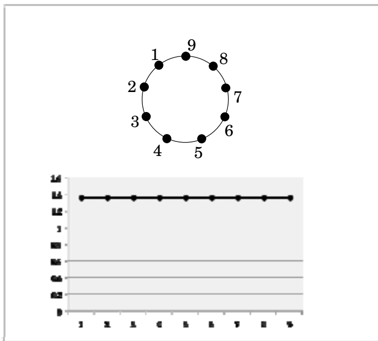


Figure 7: RRA values of two hypothetical network systems (connected through a single node resembling a bridge between two riverbanks)

5. APPLICATIONS: EXAMPLES AND CASE MATERIAL

5.1 Hypothetical Cases of Short Strings

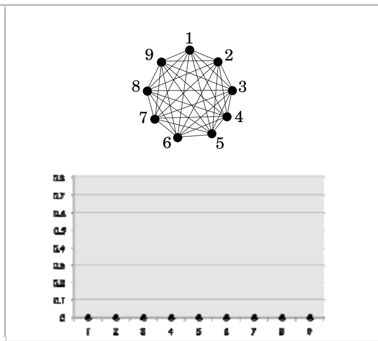
The following hypothetical cases are generated to inspect the relation between highly connected nodes and the T-complexity and T-entropy measures. The patterns vary between orderliness and structured configurations. The RRA value is utilised to search for the most integrated node(s) in each pattern and process the comparison with the string measures.



Circular Boundary System

- Each node has two links only; with the preceding and the following
- RRA values are identical for all nodes
- The “overall” entropy of the system equals 1, which is the maximum value in case of applying Shannon’s theory. This means two choice are equally probable, “link” or ‘no link”

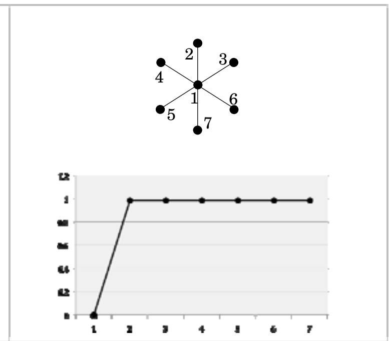
Example: Node: 9
String processed: 10000001
Length (chars): 8
t-complexity | t-information | t-entropy
3.81 (taugs) 4.5 (nats) 0.5 (nats/char)
— 6.6 (bits) 0.8 (bits/char)



Fully Linked/Saturated System

- Every node is connected with all the others in the system
- RRA values are equal and all equals zero since the system is fully saturated and symmetrical
- All the nodes deliver low entropy value since there is only one probability of choice in the system (all are “linked”)

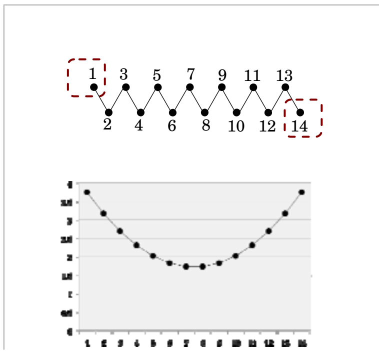
Example: Node: 1
String processed: 11111
Length (chars): 5
t-complexity | t-information | t-entropy
2.32 (taugs) 2.4 (nats) 0.4 (nats/char)
— 3.5 (bits) 0.7 (bits/char)



Radial/Polar System

- The central node is strongly connected to others on the periphery
- The peripheral nodes have one link only with the central one
- RRA values are equal for all nodes except the central one that delivers zero RRA due to its high integration within the system
- Entropy at central node 1 is low since only one choice is possible to go anywhere within the system

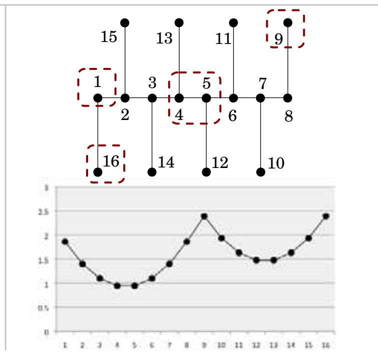
Example: Node: 1
String processed: 111111
Length (chars): 6
t-complexity | t-information | t-entropy
2.58 (taugs) 2.8 (nats) 0.4 (nats/char)
2.58 (taugs) 4.0 (bits) 0.6 (bits/char)



Saw-Tooth Sequential System

- RRA values vary along the system and take the form of “catenary” parabolic chain
- The more points towards the centre (intermediary nodes), the less RRA value. This means integration increases whenever the nodes are set in the middle of the network

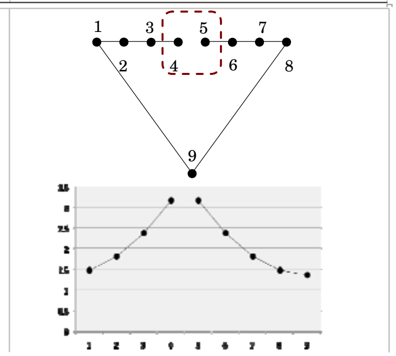
Example: Node: 8
String processed: 0000001100000
Length (chars): 13
t-complexity | t-information | t-entropy
4.95(taugs) 6.65(nats) 0.45(nats/char)
4.95(taugs) 9.65(bits) 0.7(bits/char)



Doubly-Loaded/Staggered System

- A linear route with four prongs branching out in both sides and staggered. This configuration has three levels of connections: nodes with solo link only, nodes with two links, and nodes with triple links
- Given the RRA values, the lowest values are delivered by the intermediary nodes with more links (4 & 5), and the highest are delivered by the outer-edge nodes (1,8,9,16)

Example: Node: 4
String processed: 0010100000001000
Length (chars): 16
t-complexity | t-information | t-entropy
5.38(taugs) 7.57(nats) 0.47(nats/char)
5.38(taugs) 10.92(bits) 0.68(bits/char)



Incomplete/Disconnected System

- This system represents a disconnected urban fabric where a move through a full loop is necessary to access the other side
- RRA values vary, the highest measures are delivered by the far-side nodes (on both edges)

Example: Node: 1
String processed: 10000001
Length (chars): 8
t-complexity | t-information | t-entropy
3.81 (taugs) 4.5 (nats) 0.5 (nats/char)
— 6.6 (bits) 0.8 (bits/char)

Figure 8: Values of RRA and String of Information for Hypothetical Cases of Small Systems

According to the results of short strings, the following points can be concluded at this preliminary stage:

1. The high certainty of prediction in some networks might deliver only one choice (100% choice); thus entropy equals zero if applying Shannon's equation, and T-entropy decreases if applying T-code algorithms.
2. First, the total number of relations in any system of size "n" is n(n-1)/2. However, the size of any subgraph string equals (n-1). (See Figure 6)
3. Despite the differences between the RRA values in any system, it might happen that all nodes have the same string measures since all have the same percentage choices (number of links).
4. The fall of T-complexity and T-entropy indices with the rise of RRA in the cases looked at misleads our hypothesis and causes disruption to the correlation values. The reason for this is the lack of accuracy experienced with short strings of information (less than 20 codes).
5. Either Shannon entropy or the "deterministic" entropy is "inversely" proportional with RRA, but sinceintegration equals (1/RRA), the question arises of whether this confusion comes about because of inaccurate computation of short strings, or might there be another parameter that has its effect on both measures?

The application of T-complexity and T-entropy is tricky in this sense. Two points can be made from our experience of processing the computation method: (1) The position of nodes within the system determines the synthesis (structure of symbols) of the extracted string since the connections (links) that could be made from a certain node to the other(s) are based on the choices of routes/links; (2) Since each node's "forelink" is another point's "backlink" within the system, then an introduction to some "redundancy" in this way should be considered in the estimation process to avoid replications (in case of concatenating the overall strings into one for the whole system). To reconcile these findings, another series of long string cases are analysed in the next section.

5.2 Hypothetical Cases on Large Systems

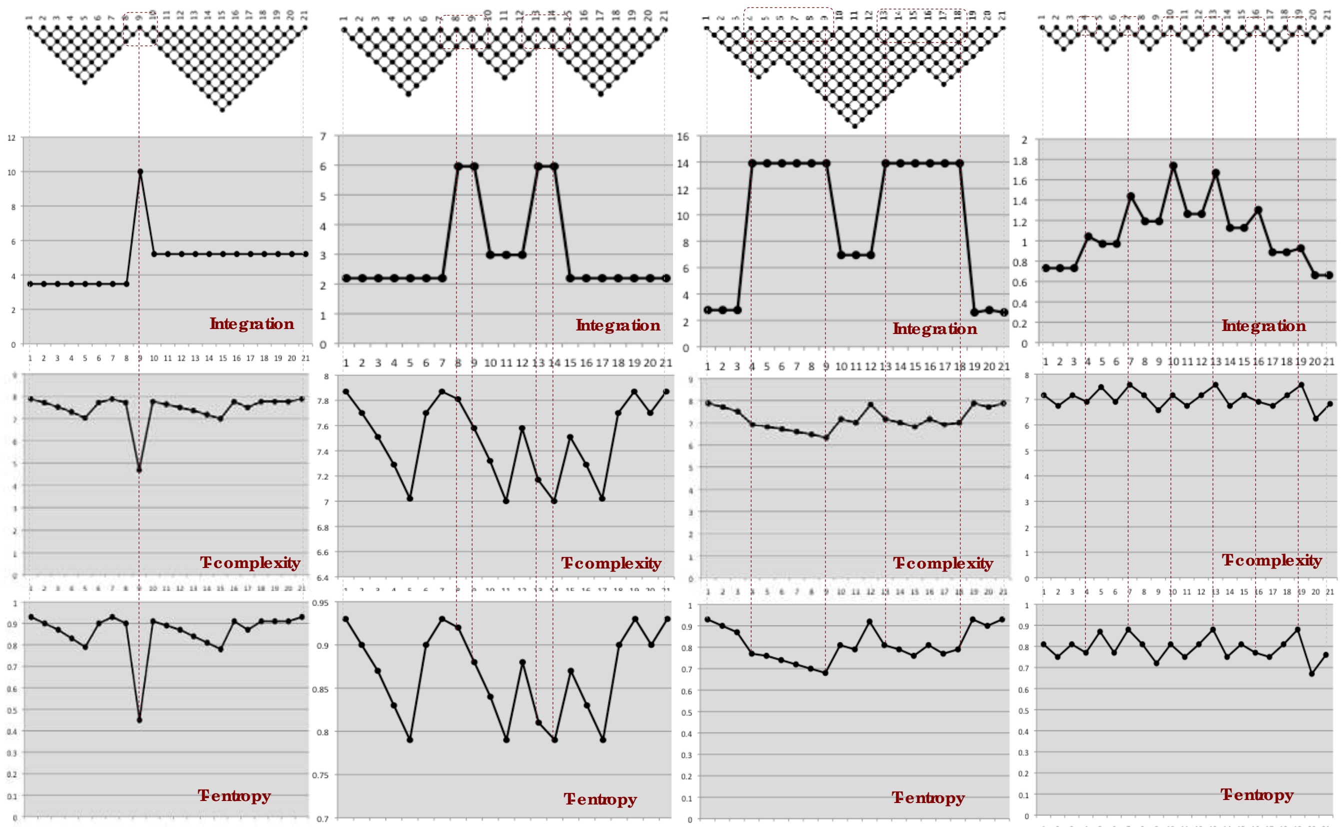
The case studies are extended to include the analyses of eight examples of longer length in order to further test the hypothesis and to overcome the inaccuracy experienced with short strings. These hypothetical systems are divided into two categories: modular order and structural, where the former is known by its repetitive and rhythmic patterns and the latter is distinguished by its variation of choices. Syntactical and string measurements are applied to study the degree of correlation between integration and "dynamic Tcomplexity" and "dynamic T-entropy".[9] See Figures 9a and 9b.

Table 2 shows the correlation values for all the hypothetical cases. According to Figure 10 (drawn from the results of Table 2), a strong inverse correlation between integration and T-complexity and T-entropy is proved in all the ordered cases. Figure 9a shows the highlighted nodes in each case. The shallower the system (e.g. case 4), the higher the degree of correlation between integration and T-complexity. The denser the system (e.g. case 3), the lower the degree of correlation between integration and T-complexity. However, in Figure 9b, only one case delivers a high correlation between integration and T-complexity, reaching 0.55 in case 3. This lack of evidence is due to the low degree of diversification in the structure of the system (the pattern is shallow) that was not the case in the other patterns. T-complexity and T-entropy are un-correlated along ordered and structured cases.

[9] The term dynamic entropy, introduced by Gero et al. (2011) to indicate that each node in the system has its own entropic measure and therefore the values fluctuate along the linkography. See also: Kan and Gero (2011a; 2011b).

Variables of Correlation	Modular Order Systems				Structured Systems			
	Case 1	Case 2	Case 3	Case 4	Case 1	Case 2	Case 3	Case 4
Integration : T-complexity	-0.76	-0.22	-0.86	0.39	0.44	0.21	0.55	-0.01
Integration : T-entropy	0.17	-0.13	0.022	-0.04	-0.21	-0.24	-0.28	0.09
T-complexity : T-entropy	0.25	0.07	0.01	0.20	0.25	0.07	0.24	0.18

Table 2: Values of correlation for the hypothetical systems



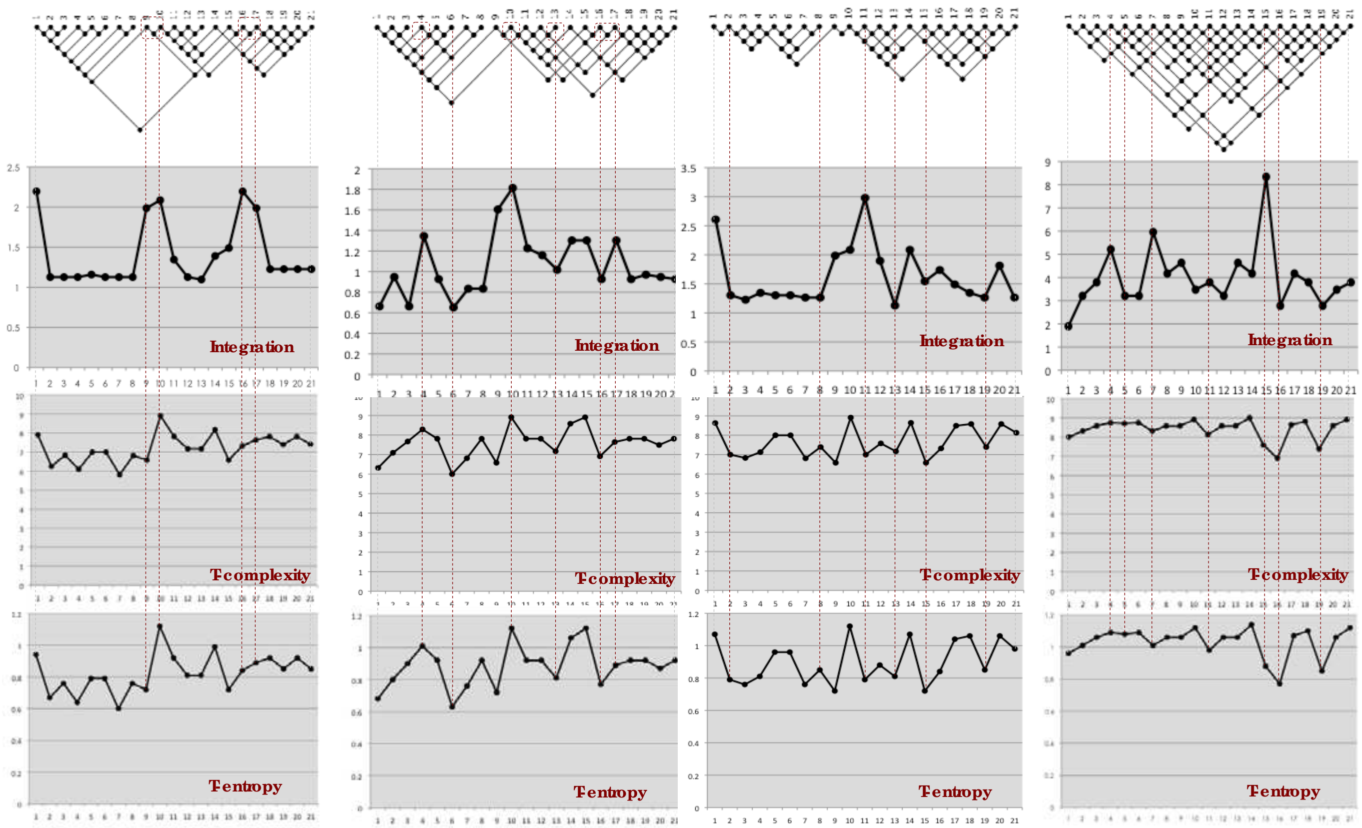
Case 1: Order Form – Two-Networks System with One Pivotal Node

Case 2: Order Form – Three Networks System with Four Pivotal Nodes

Case 3: Order Form – Three Networks System with Twelve Pivotal Nodes

Case 4: Ordered Linkography System Comprising Seven Networks – Six Pivotal Nodes

Figure 9a: The relationship between integration values and T-complexity and T-entropy in Ordered Linkography-system



Case 1: Low-Dense Structured Linkography System; Comprising Five Networks – Six Pivotal Nodes

Case 2: Semi-Dense Structured Linkography System Comprising Five Networks – Six Pivotal Nodes

Case 3: Structure Form – Diffused Networks Linkography System

Case 4: Structure Form – Highly dense Network Linkography System

Figure 9b: The Relationship Between Integration Values and T-complexity and T-entropy in Structured Linkography System

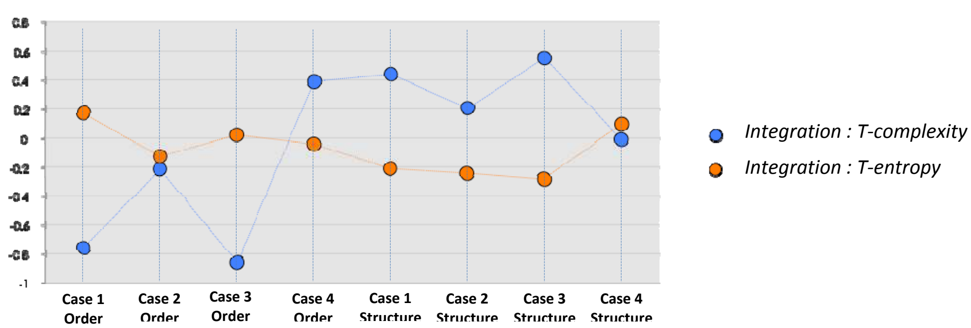


Figure 10: The Correlation Values Of Integration: T-complexity and Integration: T-entropy

According to Figure 10 (drawn from the results of Table 2), a strong inverse correlation between integration and T-complexity and T-entropy is proved in all the ordered cases. According to Table 2, Figure 10 illustrates the correlation values for each case study. It is apparent that some values are negative: "negative correlation". This means that in a relationship between the two variables one variable increases as the other decreases and vice versa. A perfect negative correlation means that the relationship that appears to exist between two variables is highly negative (might reach -1) of the time.[10] Nevertheless, the inverse correlation between T-complexity and T-entropy brings out another point to test. It is hypothesised that the more complex a string (the variety of symbols), the higher the probability of uncertainty. In short, would entropy increase with higher complexity measures? How would the hypothesis of a converse correlation with integration be affected?

5.3 Application To Real Linkographies

Figure 11 presents two different linkographies based on real design processes. The most integrated nodes are identified and correlated with T-complexity and T-entropy in addition to two other graph parameters, "betweenness" and "closeness centrality". The values of correlation are listed in Table 3. Both systems depict the following outcomes:

1. A significant correlation between integration and closeness centrality.
2. A significant correlation between T-complexity and T-entropy particularly proves the earlier result that short strings computations are inaccurate and require to be inspected through large systems.
3. A direct correlation between integration and closeness centrality.
4. An inverse correlation between integration and each of T-complexity, T-entropy and betweenness.

Variables of Correlation		Linkography 1 (size n = 328)	Linkography 2 (size n = 453)
Integration	T-complexity	0.23	0.23
Integration	T-entropy	0.22	0.23
Integration	Closeness Centrality	0.73	0.85
Integration	Betweenness	0.07	0.21
T-complexity	T-entropy	0.99	0.98
T-complexity	Closeness Centrality	0.46	0.39
T-complexity	Betweenness	0.37	0.24
T-entropy	Closeness Centrality	0.46	0.39
Closeness Centrality	Betweenness	0.37	0.41

Table 3: Values of correlation for two large linkographies

[10] A correlation in which large values of one variable are associated with small values of the other; the correlation coefficient is between 0 and -1. It is also possible that two variables may be negatively correlated in some, but not all, cases. A perfect negative correlation is represented by the value -1.00, while a 0.00 indicates no correlation and a +1.00 indicates a perfect positive correlation (definition from http://www.investopedia.com/terms/n/negative-correlation.asp#ixzz1ceYmvKXE).

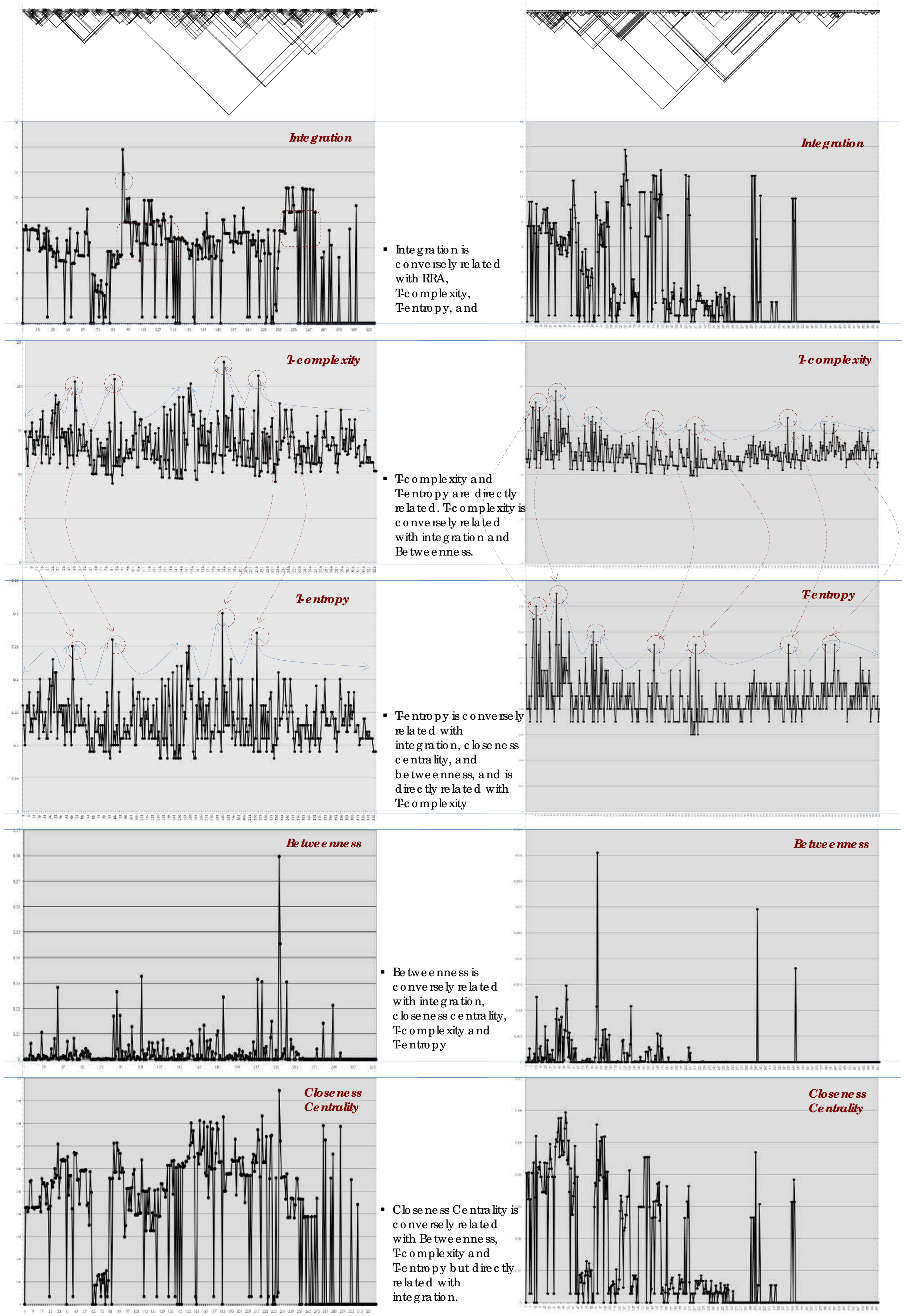


Figure 11: Different quantitative measurements for a design case study

6. IN CONCLUSION

In this paper, we have been investigating the applications of certain measures that come from space syntax analyses of urban graphs to look at linkography systems. One hypothesis is that complexity is created from the local sub-graph at different scales in the graph system than from the whole system. Since linkography and urban systems deal with multi-level complexities, the overall goal of the proposed analytical method is to reveal the relationship between the parts (sub-systems) that constitutes the system and the whole.

Two perspectives are given: the entropy theorist who looks at the overall distribution of sets of items that form the system while the information theorist looks at the individual sequence of items or the arrangement of sets that will probably occur. The application to linkography and the point depth entropy are examples of the former while the T-code computation of strings of information is adopted in this paper to look at the latter. Two different contexts are given in the case studies. Since urban configurations and linkography systems are drawn from different characteristics, the assumption is thus made to examine whether the syntactical and string parameters receive similar correlation responses in both contexts or not.

The methodology merges syntactical and string measures to highlight the significant nodes in any system and investigate the proposed hypothesis: are *highly intelligible systems associated with complexity and entropy*? Since *intelligibility, complexity*, and *entropy* are "system" properties, the method to process any system of "n" size is an aggregation of "sub-graphs" for each node in the system. The case studies include small and large systems, hypothetical and real. In order to highlight the significant nodes further, other parameters are added into the correlation: real relative asymmetry, closeness centrality and betweenness.

The relationships between string measures (T-complexity and T-entropy) and syntactical measures (integration and real relative asymmetry – RRA) are not clearly defined because of the inaccuracy of short barcodes. The assumption is then made that variable length barcode holds within it many possibilities and choices. Proving this hypothesis requires further investigation with larger systems.

The more a node is connected to the surroundings, the greater the repetition frequency in the barcodes, the less predictable the information, and therefore low string complexity results. The asymmetry of the overall distribution of nodes within the system accounts for the "associativeness" in the system and consequently gives an indication of the structure. RRA and integration values (inverse measures) can be tracked to trigger the degree of associativeness and incubation within the system.

The importance of this study lies, on one hand, from the definition it purveys about the responsiveness between the configuration of a system and the internal structure. On the other hand, it provides an analytical framework to acknowledge the degree of homogeneity between the "parts" and the "whole".

To study a configuration that underlies arrangements of nodes is about the "exposition" of facts that are called "orderly" when the observer can grasp both their overall structure and the ramifications in some details.

ACKNOWLEDGEMENT

The author wishes to acknowledge the contribution by Dr Ulrich Speidel, who has provided clear directions in the study of deterministic information theory and the computation of string measures: entropy and complexity. I am grateful for his generous advice and comments. I am indebted also to Mr. Mohamed

Abdullah for his thoughtful suggestions and technical support and to Dr Sheep Dalton for providing software to compute the integration values. Last but not least, I wish to thank Prof. Alan Penn, Prof. Ruth Conroy-Dalton and Mr. Sean Hanna for the constructive supervision, critical and inspiring throughout. The author takes full responsibility for all remaining shortcomings in this paper.

BIBLIOGRAPHY

Arnheim, R. (1971) Entropy and Art: An Essay on Disorder and Order, Berkeley, CA: University Of California Press, (2001).

Brettel, A. (2006) “The Effect of ‘Order’ and ‘Disorder’ on Human Cognitive Perception in Navigating through Urban Environments”, MSc Thesis, University College London.

Gero, J.S, Kan, J.W.T. and Pourmohamadi, M. (2011) “Analysing Design Protocols: Development of Methods and Tools”, in ICoRD'11 on Research into Design, Indian Institute of Science, Bangalore, 10–12 January 2011.

Goldschmidt, G (1990), Linkography: assessing design productivity, in R Trappl (ed) Cyberbetics and system ’90 World Scientific, Singapore, pp 291- 298.

Goldschmidt, G. (1992) “Criteria for Design Evaluation: A Process-oriented Paradigm”, in Y.E. Kalay (ed.) Evaluating and Predicting Design Performance, New York: John Wiley & Son, pp. 67–79.

Goldschmidt, G. (1995) “The Designer as a Team of One”, Design Studies, 16 (2), pp. 189–209.

Hanson, J (1989) “Order and Structure in Urban Space: A Morphological History of London”, PhD Dissertation, University College London.

Kan, J.W.T. and Gero, J.S. (2005a) “Can Entropy Indicate the Richness of Idea Generation in Team Designing?” in A. Bhatt (ed.), CAADRIA ’05, New Delhi: TVB, Vol. 1, pp. 451-57.

Kan, J.W.T. and Gero, J.S. (2005b) “Design Behaviour Measurement by Quantifying Linkography in Protocol Studies of Designing”, in J.S. Gero and U. Lindemann (eds), Human Behaviour in Designing '05, Sydney, Australia: Key Centre of Design Computing and Cognition, University of Sydney, pp. 47-58.

Kan, J.W.T and Gero, J.S. (2005c) “Entropy Measurement of Linkography in Protocol Studies of Designing”, In J.S. Gero and N. Bonnarde (eds), Studying Designers '05, Sydney, Australia: Key Centre of Design Computing and Cognition, University of Sydney, pp. 229-45.

Kan, J.W.T. and Gero, J.S. (2007) “Can an Objective Measurement of Design Protocols Reflect the Quality of a Design Outcome?” International Conference On Engineering Design, ICED ’07, Cite Des Sciences Et De L'Industrie, Paris, 28–31 August 2007.

Kan, J.W.T. and Gero, J.S. (2008) “Acquiring Information from Linkography in Protocol Studies of Designing”, in Design Studies, 29(4), pp. 315-37.

Kan, J.W.T. and Gero, J.S. (2009a) “Using the FBS Ontology to Capture Semantic Design Information in Design Protocol Studies”, in J. McDonnell and P. Lloyd (eds), About: Designing: Analysing Design Meetings, CRC Press, pp. 213-29.

Kan, J.W.T. and Gero, J.S. (2009b) "Using Entropy to Measure Design Creativity; Using a Text Based Analysis Tool on Design Protocols", in digital proceedings of the International Association of Societies of Design Research (IASDR), Coex, Seoul, Korea, 18–22 October 2009.

Kan J.W.T. and Gero, J.S. (2011a) Learning to collaborate during team designing: Quantitative measurements, in A Chakrabarti (ed), Research into Design, Research Publishing, pp. 687-694.

Kan, J.W.T. and Gero, J.S. (2011b) Comparing designing across different domains: An exploratory case study, in S Culley, B Hicks, T McAloone, T Howard and Y Reich (eds), Design Theory and Methodology, Design Society, Glasgow, pp. 2:194-203.

Kan, J.W.T., Zafer, B. and Gero, J.S. (2006) "Comparing Entropy Measures of Idea Links in Design Protocols", in Design Computing and Cognition '06, Part 4, pp. 265-84.

Kan, J.W.T., Zafer, B. and Gero, J.S. (2007) "Comparing Entropy Measures of Idea Links in Design Protocols: Linkography Entropy Measurement and Analysis of Differently Conditioned Design Sessions", Journal of Artificial Intelligence for Engineering Design, Analysis and Manufacturing (AIDAM), 21(4), pp. 367-77.

Kuntz, P.G. (1968) The Concept of Order, Seattle, WA: University of Washington Press.

Laing, R.D. (1965) The Divided Self, Baltimore, MD: Penguin.

Planck, M. (1969) Vortrage und Erinnerungen, Darmstadt, Germany: Wissenschaftliche Buchgesellschaft.

Shannon, C.E. and Weaver, W. (1949) The Mathematical Theory of Communication, Urbana, IL: University of Illinois Press.

Speidel, U. (2008) "T-code and T-information Theory", Online Tutorial, http://citr.auckland.ac.nz/~ulrich/research.html

Speidel, U., Titchener, M.R. and Yang, J. (2006) "How Well Do Practical Information Measures Estimate the Shannon Entropy", CSNDSP, Patras, Greece, 19–21 July 2006.

Titchener, M. (1998a) "Deterministic Computation of String Complexity, Information and Entropy", International Symposium on Information Theory, MIT, Boston, 16-21 August 1998.

Titchener, M. (1998b) "A Deterministic Theory of Complexity, Information and Entropy", IEEE Information Theory Workshop, San Diego, February 1998.

Titchener, M. (1998c) "A Novel Deterministic Approach to Evaluating the Entropy of Language Texts", Third International Conference on Information Theoretic Approaches to Logic, Language and Computation, Hsi-tou, Taiwan, 16–19 June 1998.

Titchener, M. (2004) "The Practical Measurement of Information and Entropy", Online tutorial: http://tcode.auckland.ac.nz/tutorial.html.

Titchener, M. (2008) "An Introduction to T-codes", Online Tutorial, http://tcode.auckland.ac.nz/~mark/T-codes%3A%20Intro.html

Titchener, M., Nicolescu, R., Staiger, L., Gulliver, A. and Speidel, U. (2005) "Deterministic Complexity and Entropy", Fundamenta Informaticae, 64(1–4), pp 443-61.

Turner, A. (2007) "Depthmap: A Program to Perform Visibility Graph Analysis", in Proceedings of 6th International Space Syntax Symposium, İstanbul, 2007.

Tversky, B, and Zacks J.M. (2001) "Event Structure in Perception and Cognition", Psychology Bulletin, 127(1), pp. 3–21.