Sketching a Network Portrait of the Humber Region

ALEXANDRA S. PENN,1,2 PAUL D. JENSEN,1 AMY WOODWARD,2 LAUREN BASSON,1,3 FRANK SCHILLER,2 AND ANGELA DRUCKMAN1

1Centre for Environmental Strategy, University of Surrey, Guildford, Surrey GU2 7XH, United Kingdom
2Department of Sociology, University of Surrey, Guildford, Surrey GU2 7XH, United Kingdom and
3Green Cape, Cape Town, South Africa 8000

Received 27 February 2014; accepted 3 March 2014

Industrial systems can be represented as networks of organizations connected by flows of materials, energy, and money. This network context may produce unexpected consequences in response to policy intervention, so improved understanding is vital; however, industrial network data are commonly unavailable publically. Using a case study in the Humber region, UK, we present a novel methodology of “network coding” of semistructured interviews with key industrial and political stakeholders, in combination with an “industrial taxonomy” of network archetypes developed to construct an approximation of the region’s networks when data are incomplete. This article describes our methodology and presents the resulting network. © 2014 Wiley Periodicals, Inc. Complexity 000: 00–00, 2014

Key Words: industrial ecology; industrial networks; network analysis; incomplete data; industrial network archetype

INTRODUCTION

The use of the tools and techniques of complexity science to support decision making in real world scenarios has significant potential but encounters challenges specific to its social context. This article presents work developed on a project that aims to use complexity science to develop models that aid policy makers and regional stakeholders in decision making about their industrial “ecosystems.” That is, the systems formed by companies and other organizations in a particular locale, which are connected by supply chains, waste streams, power or heat supply, and other nonmaterial relationships, such as service provision.

Complexity science, in particular network analysis, is considered to be an appropriate and potentially useful tool to explore such industrial ecosystems or networks because of the interconnected nature of the industrial actors present. This interconnectedness gives rise to the possibility of indirect or “emergent” system-level effects when certain policy instruments or combinations of such instruments are applied. Such effects may be out of the range of experience or intuition of the decision makers involved (or the current modeling methodologies they

Correspondence to: Alexandra S. Penn, Centre for Environmental Strategy, University of Surrey, Guildford, GU2 7XH, United Kingdom. E-mail: a.penn@surrey.ac.uk

©2014 Wiley Periodicals, Inc., Vol. 00 No. 00
DOI 10.1002/cplx.21519
Published online 00 Month 2014 in Wiley Online Library (wileyonlinelibrary.com)
might potentially apply). Our intent is to provide “thinking tools” that could allow scenario exploration and the discovery of key points of intervention for policy or management. Simulation modeling of the particular industrial ecosystem in question would be the ideal methodology to use for this goal; however, such models require the availability of detailed or specific information on the system in question.

The presented research derives directly from this project and the challenges that were faced as it progressed. Real world contexts, such as industrial ecosystems, are populated by numerous, often strategic, actors of different types, and data may commonly be sparse, commercially sensitive, or simply unavailable. These are systems of high complexity with important social elements meaning that situations are nuanced, reflexive, and potentially not straightforwardly describable in low-level causal terms. From this context, this work represents a specific methodological response to the problems of extracting and developing network models from the incomplete and often qualitative data that is available.

**Background**

**Industrial Ecology**

In the field of “industrial ecology,” the systems (or networks) formed by the complex interactions and interdependencies between firms in a given locale or within an industrial sector are regarded as “industrial ecosystems” (e.g., [1–10]). It is considered that the functioning of these ecosystems may be improved in regards to their material sustainability by emulating certain properties of biological ecosystems perceived as desirable, such as high nutrient cycling ratios. While there are many different, potential conflicting, definitions of sustainability, industrial ecology principally considers the promotion of so-called “industrial symbiosis,” that is the reuse or recycling of one or more companies’ by-products or waste/excess utilities as primary resources within the industrial network [11–13]. An industrial ecosystem constitutes not only a metabolic network formed by flows of energy and materials but also formal and informal social networks with economic, contractual, and social relationships through which information or money may be considered to flow (e.g., [14–18]).

The structure, nature, and development of each of these different networks are interconnected and influence that of the others. Thus, if we wish to understand the dynamics of an industrial ecosystem, we must understand the interplay of these different networks and map their individual structures. Ultimately, macroscopic properties of the industrial ecosystem that we are interested in managing, such as carbon dioxide (CO2) emissions, material throughput, regional Gross Value Added, or job creation will be the result of these structures and their lower level dynamics under the influence of regulation and other external factors.

**Network Representations**

The representation of sets of interactions or relationships between interacting entities as a network or graph has become widespread in numerous fields [19, 20]. Discrete entities, which may be individuals, organisms, species, genes, companies, etc. depending on context, are described as nodes or vertices connected by directed or undirected links or edges, which indicate the existence of a pairwise interaction or connection. Metabolic networks, such as food webs and trade networks (among others), are connected by directed links, indicating the direction in which material, energy, or money flows. Social networks, in which connections commonly indicate relationships or acquaintance, often consist of nodes connected by undirected links (although directed graphs are used in the case of asymmetrical relationships). Network analysis has proved to be a useful tool in understanding whether specific network structures are vulnerable to failure and which particular nodes in a given network exert a strong influence on its processes [21]. Similar analysis could prove useful to identify points of intervention and systemic risks for decision makers in industrial ecosystems if a sufficiently structurally accurate network can be constructed.

A large body of literature exists in the industrial ecology community considering the development and properties of interfirm collaboration and networks engaging in industrial symbiosis (e.g., [22–27]). These studies were mainly qualitative in nature; however, some more quantitative investigations into the development of industrial networks do exist. Recently, tools and metrics from network analysis have been used to explore the evolution of industrial symbiosis within particular contexts: for example, Paquin and Howard Grenville analyzed the formal brokerage of symbiosis within the National Industrial Symbiosis Programme (NISP) of the United Kingdom [28,29], while the characteristics of social network structure leading to the formation of symbioses have been studied by Ashton [9,17,18]. These example studies have largely mapped the IS network in general terms, rather than considered the explicit metabolic details of particular exchanges. Jensen et al. considered the individual physical movement of a range of materials within the industrial symbiosis network facilitated by NISP [12] but did not perform analysis of the network generated in terms of network-specific properties. All of this work focused only on the component of the regional industrial metabolism that consisted of exchanges classified as industrial symbiosis. Neither conventional supply chain links nor waste flows organized via other organizations or means were included. If we wish to model the development of a given...
Incomplete Network Data

Several approaches already exist which aim to address the problem of incomplete network data, principally within social network analysis. Essentially all of these approaches calculate certain local properties of a network subsample and use them as estimates for those properties on the whole network. A complete network can then be simulated by stochastically adding links using a variety of methodologies. One class of models which might be used here are “configuration” or “stub” models. Such models are often used to create random networks with the same degree distribution as a given network by breaking up a network into its constituent nodes and their half-links or “stubs” and randomly reassembling them [30,31]. A similar technique can also be used to generate a random network or ensemble of networks with the same degree distribution as the sampled network. In this case, we would create a set of nodes \( n \) with numbers of stubs, based on the degree distribution calculated from the subsample. The network (or family of networks) can then be completed by randomly connecting these half edges. Node types can additionally be randomly assigned to nodes in the networks thus generated based on the composition of the subsample, potentially including information on correlation of degree with node type (differential degree). Clearly in such a model, higher order structural information from the network subsample, such as which node types are likely to connect with others, or properties such as transitivity and clustering, will not be preserved as only node-level information is considered.

Many methodologies for incomplete networks fall within the framework of Exponential Random Graph Models (or ERGMs). An ERGM is the set of all possible randomly constructed graphs with number of nodes \( n \), which can be modeled by a probability distribution of a given number of local, node-level configurations of the measured network [32]. Configurations consist of node-link patterns or motifs, such as edges, triangles, and stars, within nodes’ ego networks. Originally, the parameters of the probability distribution of configurations used to generate an ERGM family were calculated, using Maximum Likelihood Estimation techniques, from what were assumed to be complete network data. In this case, the random networks generated are used for testing hypotheses on the formation or general properties of networks of a particular structural type. However, ERGMs can also be used to construct networks from a subsample of data [33–37]. In these situations, the models used to randomly simulate complete networks are again based on the specific local properties of the subsample [38]; in this case, the distribution of node-level configurations within the sample. Depending on the particular configurations taken into account in the construction of the probability distribution, more complex structural information may be retained or approximated (in the case of graph completion) in the graphs produced, for example, reciprocity, homophily, transitivity, and clustering.

There are several reasons why neither of these broad classes of method is suitable for our data or for the uses to which we wish to put our constructed industrial network. First, we do not have complete connection data for the nodes which we have interviewed, rather we have dense, but incomplete, local data on our interviewed nodes plus a great deal of sparse detail on other nodes in the network. Hence, we do not have a good estimate of node-level properties with which to construct a full network. A configuration model would require reliable information on degree at the node level in order to produce a good approximation of the degree distribution of the whole network. The calculation of the probability distribution of configurations used to create an ERGM requires complete ego network data (that is a focal node, all of its first degree connections, and the connections between those neighbors) for the nodes sampled. We have neither. Second, our sampled nodes are neither independent nor representative of the whole network. Our nodes are not sampled independently, as data were necessarily acquired through snowball sampling in this hard to reach population. Additionally, the characteristic community composition of this area is highly diverse, with many different industry types. While some sectors contain numerous firms, many others contain very few. We know that there are only a small number of nodes of certain regionally important types. For example, the area supports one cement work and two oil refineries, none of which we were able to interview. It does not seem justifiable to
assume that these diverse industrial sectors could be adequately represented by the metabolic network connection structures of other very different types given how strongly constrained they are by physical and chemical realities. Sampling error would, thus, be very large using either of the above approaches. Third, all of these methods construct networks by assigning connections between nodes stochastically. However, the specifics of who connects to whom matters for the uses to which we wish to put this model. In the main, we are not interested in statistical properties of a network of this type, rather we wish to produce a structure which maps sufficiently to reality so that we may explore the impact of particular shocks, stresses, or changes to the network on particular sectors or to identify key industry types on which to focus policy interventions. Within these relatively small and diverse industrial networks, node types are highly constrained in who they would connect to and network behavior is likely to be highly contingent on this specific structure. Therefore, randomly generated networks will not be sufficient for our purposes and would not provide the insight we need to inform local decision making.

In the absence of complete information, we, nevertheless, have access to expert information on the characteristic inputs and outputs of different types of industry, both generally from a chemical engineering perspective and more particularly in the regional context from site-specific research. We propose that a more useful network representation can be obtained by incorporating this expert knowledge into the specific data obtained from interviews. In light of our specific data availability and requirements for our model, the development of a novel methodology for constructing sector-specific network archetypes is proposed.

**Generation of Industrial Network Archetypes**

Configurations or network motifs, that is, specific subgraphs consisting of characteristic connection patterns between sets of vertices, have been increasingly discussed in a variety of network contexts [39]. Configurations are used to construct random networks within an ERGM methodology as described above, while a related concept, motif analysis, has been used to analyze numerous large datasets. Commonly in motif analysis, the representation (number) of each of the all possible motifs for a given number of vertices in a particular network is calculated and compared to the expected number of each motif present in a randomly connected graph with the same number of vertices and edges and with degree properties preserved at the individual node level. Motif analysis has been performed on numerous types of biological networks, and suggestions have been made that some significantly over-represented motifs correspond to particular functions that are selectively retained over evolutionary time [40–42]. More recently, motif analysis has been performed on interfirm trade networks such as Ohnishi et al.’s study of material and service transactions in the Japanese economy [43].

All of these analyses were made possible by the existence of large and relatively complete network datasets, not available in our situation. As previously discussed, within the constraints of this project and much applied complexity science, a full industrial network is not practically obtainable, but our goal is to provide a good enough model to explore and capture qualitative network dynamics. In this article, therefore, we take a different approach constructing characteristic connection patterns for individual industries based on interview data, extra information from documentary sources, and expert first-hand knowledge derived from working in the case study region. It should be noted that the motifs that we construct will have variable numbers of vertices and be unique to each industry type; for this reason, to avoid confusion, we will use the term network archetypes to describe them. This approach was inspired in part by work in economic geography in which different categories of industrial districts, characterized by different interfirm supply chain network structures and abundance of different firm types, have been identified (see e.g., [44]). Different district structures are thought to correspond to differential economic performance as well as resilience and other macroscopic properties and to require different governance or intervention strategies. Rather than considering characteristic connection structures of a whole district, however, we consider an approach inspired by functional ecology in which industries will have different connection structures based on their economic and metabolic functions. Just as species are commonly categorized as, for example, generalist or specialist predators by the number and diversity of their connections in a trophic network structure (see, e.g., [45]), we have attempted to produce a simplified classification of industries via their metabolic network connections. From this perspective, the industrial network archetypes explicitly correspond to functional types, which we may think of as a company’s “niche” in the industrial ecosystem. It is hoped that “plugging in” these archetypes into the incomplete network coded from the interviews will allow us to structure the network more realistically.

**DEVELOPMENT OF METHODOLOGY**

In the sections below, we describe the methodological development in sequential steps: from stakeholder engagement and qualitative interviewing; to the initial network coding from qualitative data; to the development of industry-specific metabolic network archetypes, through to the elaboration of the network graph. We describe the
method used in each step, show and discuss the results that it can produce in our context, and explain how they have led us on to develop the next steps.

Case Study
The particular industrial ecosystem chosen for our case study is the network of industries situated in the area surrounding the Humber Estuary in the United Kingdom. The focus of the work is to understand how various policy instruments, including efforts to promote resource reuse via industrial symbiosis, affect system structure and dynamics. The Humber region is a large, active industrial area comprising a diverse set of industries ranging from the United Kingdom’s highest concentration of food processing industries to major oil refining and chemical and biochemical production facilities. The ports of Immingham, Grimsby, Goole, and Hull, form one of the largest and busiest port complexes in Europe. The estuary provides infrastructure for 20% of national gas landing and 27% of UK oil refining capacity [46] and the wider region is the source of 27% of total UK CO₂ emissions [47].

The estuary is of national and international biodiversity and conservation importance and due to climate change presents increasing flood risk management issues, both of which issues can cause friction over proposed development. Neighboring communities face significant socio-economic problems including unemployment and fuel poverty. Development of the region is affected by, and affects, linked biophysical, industrial, economic, social, and governance systems, populated by many diverse actors. The region faces significant new challenges and opportunities with transition to a low carbon economy and national energy security as current key and potentially controversial policy issues. It is one of the most important energy hubs of the United Kingdom, with strategic energy generation facilities and infrastructure, significant potential for carbon capture and storage, and new investment in large-scale renewable energy technologies from offshore wind to biofuels. Decision making about the region and its possible future scenarios will have impacts on sustainability goals locally, nationally, and globally.

The region is also significant for its role as an early area of focus for the industrial ecosystem concept in the United Kingdom. (The creation of more sustainable industrial ecosystems with increased efficiency in resource use is often attempted via the earlier discussed practice of industrial symbiosis.) In 2005, the UK Government’s Department of Environment, Food and Rural Affairs funded the development of NISP [48]. A national, but locally delivered, program aimed at providing the collective expertise and networking resources needed to facilitate resource exchanges between industries within various industrial areas around the United Kingdom. Extensive NISP archival data have in the past been used by others to explore industrial symbiosis in the Humber region and nationally within the United Kingdom (e.g., [12,49,50]). However, for this study, these commercially sensitive data were not accessible.

Semistructured Interviews
This article reports on the data and analysis gathered from interviews conducted with regional and national stakeholders recruited via a snowball sampling strategy. Our initial engagement with stakeholders in the region revealed that data were not easily accessed due to their commercial sensitivity and the lack of relevant publicly available datasets. Additionally, the region is characterized by satellite industry. That is, branches of large multinational organizations with headquarters or principal decision makers located elsewhere. Given this context, snowball sampling was identified as a suitable method with which to access participants’ social networks in order to reach specific populations which would have otherwise been hidden or inaccessible [51,52]. Gatekeeper interviewees, who subsequently opened up their personal networks for our utilization, were initially identified and enlisted through repeated attendance at regional forums over a number of months and presentations in which the goals and potential approach of the project were introduced. This fostering of trust and respect with participants was subsequently communicated to specific stakeholders in their network either through introductions at regional events or direct invitation to participate in interviews or workshops we held in the region.

While utilizing a nonprobability sampling technique such as snowball sampling has constraints, such as the bias of the sampled population, and the representativeness of results [53], utilization of the method enabled the completion of 18 semistructured interviews with key regional industrial and political stakeholders over a 7-month period (see Table 1 for details of company/organization types interviewed), which would not have otherwise been possible.

The interviews were conducted predominantly face-to-face, with two taking place over the telephone due to time constraints of participants. Face-to-face interviews were favored due to the semistructured format, allowing for more in-depth interviewing. The rapport developed through face-to-face communication is vital [54], particularly when questions can be of a sensitive or contentious nature as in this project. A semistructured interview script was designed in relation to the overall objectives of the project, the need for the interview data to “feed into” multiple subprojects, and the information accessible by the targeted participants. Interviews focused on the “symbiotic” or other connections that companies were engaged in or had explored as possibilities: the ways in which the need or opportunity for connection had arisen; the means
by which partner organizations were found; the nature of the decision-making process; barriers and facilitating factors to both creating and maintaining links; the consequences of such link formation and the situations in which links might be discontinued. Importantly, questions covered the social as well as technical, economic, and regulatory influences on the nature and dynamics of connections. Additional questions covered the development of the regional network as a whole over time, expected future development, and aspirations.

The utilization of a semistructured method not only allows for some comparability between interviews but also ensures that additional factors not designed for, individual experiences and perspectives, can emerge and be elaborated upon during the process [55]. These qualitative interviews were designed to take place with individuals and to last between 30 min and 1 h though on several occasions none of these designs were met. A pilot interview with an organization from outside the region, also engaging in industrial symbiosis, was conducted in order to ensure the suitability of the design.

**Network Coding**

The information gathered in the interviews contained substantial amounts of information about the networks present. In order to develop an understanding and picture of these networks, a process of coding the interviews was embarked on. Coding involves a process of identifying and symbolically labeling a text with “tags or labels for assigning units of meaning to the descriptive or inferential information compiled during a study” [56]. There are no specific guidelines for coding qualitative interviews; it is dependent on many independent factors, for example, the amount of data collected or the questions being explored. We used a mixed coding strategy: an initial code specifica-
tion was identified via “theory-led” strategy, utilizing the knowledge, we already had about the region and identifying the factors necessary for the incorporation of network analysis [57]. Simultaneously, a “data-driven” strategy was endorsed [57], supporting the emergence of codes from the text itself. Many types of labels can exist; within this project, we labeled nodes, links, and flows. Nodes can refer to individual companies or organizations, including references to specific named companies/organizations, to nonspecified instances of certain types of organization and also to potential companies/organizations of certain types that could become part of the regional network. While links and flows refer to the edges or connections between one company/organization and another or others along which materials, energy, money, or information might travel. In some cases, merely the presence or potential/former presence of a particular link of given type was mentioned, in others, more detailed quantitative or qualitative information was given in addition.

The aim of this coding process was to find repetitive patterns with which to analyze the Humber network. A crucial aspect of analysis in itself [58], coding is a heuristic in which one can explore and categorize data. The process of developing the final categorization of data, a coding scheme or coding tree [58], is a cyclical process of defining codes and assigning the codes to categories [59]. Interviews were transcribed and uploaded to Computer-Assisted Qualitative Data Analysis Software (CAQDAS) to be labeled with network-relevant information such as: node and link types, network organizations, and general comments on structure, reconfiguration, and network dynamics. Text was coded using the CAQDAS software MAXQDA platform (version 10).

The coding process starts during data collection. An initial skeletal coding tree was created containing top-level node-type and link-type categories corresponding to types of entities and connections that we knew to be present in the system from initial exploratory discussions with key regional stakeholders. These pertained, for example, to regulatory and industrial nodes in certain sectors and informational, social and material, and energetic links. As coding proceeded, new node and link types were added as subcategories, and new top-level categories were created where the need became apparent. Several passes were made through the interviews to reconsolidate and rationalize the coding tree.

---

**TABLE 1**

List of Sector and Company Types for Interviewed Companies and Organizations

| 1. Biodiesel Producer: Waste Feedstock | 10. Local Authority |
| 2. Bioethanol Producer: Virgin Feedstock | 11. Local Authority |
| 7. Facilitator/Broker (Industrial Symbiosis) | 16. Port Authority |
Text segments which mentioned particular connections of a given link type, to and from given node types, were coded with the codes for both the node and the link types in question. In situations in which nodes performed two distinct functions, for example, as an on-site energy generator and a cement works, they were coded as both. The list of the resulting link and node-type codes is given in Tables 2 and 3.

Other information was coded where available. For example, information on the links where present, such as durations, quantities, frequencies, and prices, basic node attributes and processes performed, and any mention of factors influencing connection decisions. This information was used to support the construction of industrial network archetypes (see below). Additionally, sections mentioning other factors influencing network evolution or dynamics, exogenous stresses or shocks impacting on the network and other factors potentially influencing network structure were also coded but are not made use of in this analysis. The outcome of interviews did not contain exhaustive information about the companies’ network connections. However, explicit reference to connections, either specifically to other companies, or more generally to other sectors, emerged through the interview process. These references provide the possibility of extracting a regional network, albeit incomplete.

RESULTS: NETWORK VISUALIZATION FROM INTERVIEWS ONLY

This incomplete interview network was visualized using the open-source plugin NodeXL, which provides graph drawing and analysis facilities for Excel. In this initial visualization, the network was drawn with individual nodes consolidated by sector, or "node type," to give an overview of the functional form of the metabolism. That is, in the graphs displayed here, the links to or from multiple instances of different node types are consolidated into one representative node of that type. Links are also consolidated in this visualization so that numerous links of the same type between the same two node types are compressed into one link.

Different link types are represented with different colors. In the higher level network visualizations below, dark green is virgin biomass, lime green shows bio-based products, and olive green is biological wastes. Black shows inorganic virgin resources, pink inorganic products, and orange inorganic wastes. Brown is mixed waste. Red is heat, blue electricity. Nodes include those already mentioned (see Table 2) plus nodes coded for sources and sinks in import and export of resources and energy. Links are also coded by a time label of either past (discontinued), present, planned (imminent within next few years), or potential (possible, but no plans at present), which allows us to gain an idea of the system’s possible evolution.

Figure 1 depicts all edges mentioned in the interviews, including links which are present now, potential and planned links and those which have been discontinued or for which planning permission was refused. The graph was coded from 826 coded “node” segments and 279 coded “link” segments and shows the metabolic network with resources, wastes, products, and energy only. These preliminary results allowed us to visualize more clearly

<table>
<thead>
<tr>
<th>TABLE 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Table of Metabolic Node-Type Codes from Interviews</td>
</tr>
<tr>
<td>Aggregator</td>
</tr>
<tr>
<td>Agriculture</td>
</tr>
<tr>
<td>Airport</td>
</tr>
<tr>
<td>Anaerobic Digester</td>
</tr>
<tr>
<td>Biodiesel Production:</td>
</tr>
<tr>
<td>Virgin Feedstock</td>
</tr>
<tr>
<td>Biodiesel Production: Waste</td>
</tr>
<tr>
<td>Bioethanol Production:</td>
</tr>
<tr>
<td>Virgin Feedstock</td>
</tr>
<tr>
<td>Bio-Fuel Producer General</td>
</tr>
<tr>
<td>Bio-Gas/Electricity</td>
</tr>
<tr>
<td>Biological Heavy</td>
</tr>
<tr>
<td>Intermediate Compound</td>
</tr>
<tr>
<td>Bio-Processor</td>
</tr>
<tr>
<td>Composter</td>
</tr>
<tr>
<td>Construction and Related</td>
</tr>
<tr>
<td>Export: Resources</td>
</tr>
<tr>
<td>Extractive: Aggregates</td>
</tr>
<tr>
<td>Extractive: Coal</td>
</tr>
<tr>
<td>Extractive: Limestone</td>
</tr>
<tr>
<td>and Lime</td>
</tr>
<tr>
<td>Extractive: Oil</td>
</tr>
<tr>
<td>Facilitator/broker-agent</td>
</tr>
<tr>
<td>Food Producer/Retailer</td>
</tr>
<tr>
<td>Housing</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>TABLE 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Table of Metabolic Link-Type Codes (Top-Level Coding)</td>
</tr>
<tr>
<td>Products:</td>
</tr>
<tr>
<td>Inorganic</td>
</tr>
<tr>
<td>Bio-based</td>
</tr>
<tr>
<td>Virgin Materials:</td>
</tr>
<tr>
<td>Biological</td>
</tr>
<tr>
<td>Inorganic</td>
</tr>
</tbody>
</table>
and put into context some of the information gained from our interviews.

A simple quantitative analysis of the network consisting of all links mentioned in interviews, and its vertices was performed in Gephi 0.8.2 Beta (http://gephi.org/). For the purpose of this analysis, duplicate links were deleted in order to reduce sampling error in the coding resulting from repeated mentions of the same links by interviewees. Results are shown in Table 4. As expected, the network displays signs of sampling bias probably due to the interview process. If we rank nodes by the sum of their in and out-degrees (total number of incoming and outgoing links respectively), we see that six of 10 of the (metabolic) companies that we interviewed are in the top 10 of the 50 node-types in the network, whilst other top ranking nodes either provide feedstocks for these companies or are highly general nodes such as import or export of resources (nodes with social connections only are excluded from this network although their interviews have provided data for it.). Eight out of 10 of our interviewed companies have a total number of connections above the network average.

Another way of comparing the importance of the nodes within the network is to consider their betweenness centrality [60]: a measure of the total number of times that a node acts as an intermediary along the shortest path between two other nodes in the network. If we rank all nodes in this network by their betweenness centrality, we again see evidence of potential sampling bias. Interviewees make up six out of 10 in the top 10 nodes by this measure (Gephi uses the algorithm described in [61]. Shortest paths are calculated taking directionality into account.). Clearly, nodes corresponding to companies which we have not interviewed probably have more connections than are visible here and may, thus, be more central. Indeed even for interviewed company types, due to the limitations of qualitative, semistructured interview methodology, the network graph is conspicuously lacking certain links that we know from metabolic or economic considerations must certainly be present. This may be because the interview conversation simply did not touch on them or because they were thought too obvious to even mention. Actors
may also be strategic and wish to promote certain perspectives on their organizations, or an extensive discussion of a particular topic may reveal a higher level of detail, and, thus, more network connections than areas which are in fact similarly rich, but were not as deeply explored.

To put it succinctly, the nodes representing companies that we have interviewed, and those with which they frequently interact, tend to be relatively more connected than may be realistic. The network visualizations suggest that the original network approach has the advantage of being able to extract network data from qualitative information to generate a picture of the system that is immediately compelling and which may be useful in stakeholder engagement. However, the clear evidence of sampling bias shows that the network structure is, unsurprisingly, likely to be incomplete. As extensively discussed above, this incomplete interview coverage of regional industries, the biased nature of our data and the regional industrial context render standard statistical graph-completion methods ineffective. The extracted interview network or any simulated random network based upon it would, thus, not be a suitable basis for dynamical or scenario modeling. Although with continued engagement we may obtain information with which to produce a more complete network, data will still be limited by the subset of stakeholders who can be reached via snowballing of contacts, their willingness to divulge information, and on-going research time and resources. For this reason, it was decided to fill network gaps using an approach based on constructing sector-specific industrial network archetypes.

### Creation of Industrial Network Archetypes

As described briefly earlier, our industrial archetypes are characteristic sets of input and output connections for a given industrial type, including not just the number of connections but the identity of the vertices to which they connect. The archetypes presented here incorporate material and energy flows, although water is not included. In contrast to most ecological network analysis, we have categorized our link types (see Table 3) rather than simply identify them as flows of material/energy. Crucially, flows of waste are also included. However, in this preliminary analysis we show only link presence and do not make use of link type. Characteristic connection patterns for all nodes produced from the coding process as well as other industries/metabolic nodes known to be present in the area were generated, 76 in total. As well as the types of companies that a node is likely to connect to, the archetypes also contain information about the size of those companies as this is likely to affect their decision rules.

Some of the varieties of different functional forms that exist are visible in Figure 2. For example:

1. A cement works (a branch of a multinational) has a large requirement for waste to use as fuel in its
kiln and will buy or take feedstock from almost any local supplier whatever the size (although small suppliers will usually provide materials via an intermediary waste aggregator). These are mutualistic relationships as they allow resource donor companies to avoid landfill tax. It exports its product, but will compete with other locally based companies large and small for bio-based waste products.

2. An oil refinery will have very few local suppliers of primary resources, excepting bioethanol and biodiesel for blending and some chemical catalysts. It will have many downstream relationships with a...
multitude of sectors due to the diversity and chemical complexity of its products. It may cooperate or compete with other local refineries, e.g. by building a shared combined heat and power plant or for space and grid access.

3. Food processors will take in material from a variety of sources via aggregators, as well as final products which may go to a variety of distributors and retailers of different sizes, they will feed into secondary food producers. A large variety of waste products may go into biodiesel production, biological heavy intermediates, the cement kiln, incineration in an energy-from-waste facility or into a regionally based anaerobic digester, or to effluent treatment.

4. An offshore wind installation will have negligible metabolic connections from other industries in the area once installed. Its main connection will be direct export to grid (via an onshore substation). It can be considered as a source of energy within the system, but it has very few local metabolic connections.

In order to add the archetypes to our interview coding network, they were broken down into pairwise interactions and added to a matrix, summarizing all connections with sector names corresponding exactly to node types that we had previously identified in coding (Figure 3). Again, duplicate edges were deleted. New node types were added where found to be appropriate to complete the graph. The archetypes represent a characteristic version of a single instance of a company in a particular sector, but in our current network, nodes are consolidated by sector. That is, the connections of multiple companies of the same type

![Matrix summarizing all motif connections.](https://example.com/figure3_matrix.png)

© 2014 Wiley Periodicals, Inc.  DOI 10.1002/cplx
are consolidated into the same node. For most industries, a sector-level motif and a company-level motif will be very similar, but for some industry types, such as product aggregators for example, a sector-level motif may differ substantially. An individual company may specialize in aggregating a particular product, but a sector-level aggregator motif will show connections to and from numerous sectors. We must be aware of such effects when interpreting the analysis of our network. We imported the connections matrix into NodeXL as a list of edges and added it to the interview network, again deleting duplicate edges.

The Augmented Industrial Network

The full graph with all archetypes incorporated into the network drawn from the interview data is shown in Figure 4. Again, a simple network analysis was carried out in Gephi to calculate degree, in-degree, and out-degree and betweenness centrality for all nodes.

Comparison of the Networks

In order to determine whether our addition of network archetypes has reduced sampling bias and significantly improved the structure of the graph, we can compare the results of analysis on the two networks in a variety of ways. We focus on the degree, which we have directly affected with our methodology, and the betweenness, which it changes indirectly, but which is of significant interest for network behavior.

First, we consider the form of the degree distributions of both networks; comparing in-degree, out-degree, and total degree (Figure 5). A visual inspection of the distributions suggests that the original interview network is substantially skewed toward nodes with few incoming or outgoing links, whereas the network augmented with industrial archetypes has a more uniform distribution with increased maximum values. We can check this assertion by calculating the Gini coefficient of the distributions in question. The Gini coefficient is a usually used to provide a normalized indicator of wealth inequality within a frequency distribution [62] (a value of zero represents complete equality whereas a value of one represents complete inequality). It can, however, be used to give an estimate of the skew of a distribution which is comparable between datasets from any domain (e.g., [63]). Calculating Gini coefficients (shown in Table 5), we
can see a consistent decrease. This implies that all degree distributions have been made consistently and substantially less skewed by the addition of network archetypes. As a highly skewed network is a likely consequence of our biased sample and unlikely to represent the underlying network structure, this suggests that our methodology has reduced the bias to some extent.

Returning to the ranking of industrial types within the networks, the augmented network shows less sign of bias. Only four of 10 interviewed sectors are in the top 10 for degree, and only three out of 10 are in the top 10 for betweenness centrality. We can quantify this observation more effectively by calculating the Spearman’s rank correlation coefficient for the position of the nodes present in the original interview network in both networks (Table 6). Both sets of ranks display a small negative correlation. This is consistent with what we would hope for from our methodology; that is, that relative ranking of interviewed types tends to decrease as additional connections are added via the archetypes but that overall the augmented
network has a significantly different structure to the interview network.

**A PORTRAIT OF THE HUMBER INDUSTRIAL METABOLISM**

Although we still cannot rule out sampling bias completely, the form of the network does correspond with commonly made assertions about the region. The top 10 sectors for degree, in-degree, and out-degree and betweenness centrality are shown in Table 7 and the specific results in the context of the Humber network are discussed below.

Calculating the different types of centrality of the different nodes present in the network gives us some clue as to the regional metabolic structure and its properties, as well as the relative importance of different node types. Degree, in-degree and out-degree centralities tell us how relatively highly connected nodes are. High degree is often assumed to indicate a powerful position within a network, with more opportunity to influence other nodes and more opportunity to choose between potential trading partners. In our directed network, in and out-degrees are potentially more informative however. In social network terms, nodes with high in-degree are said to be highly prominent, assuming that numerous individuals wish to connect to them, and those with high out-degree are considered to be highly influential, in that they may transmit information to or trade with many other nodes [64]. As described above, betweenness centrality measures of the total number of times that a node acts as an intermediary along the shortest path between two other nodes in the network and thus indicates its importance as a network “bridge.”

This measure was originally conceived as a way to determine the degree of control that an individual might have over the flow of information in a social network [60].

In a metabolic network context, however, these measures may imply different properties. A high in-degree may indicate a requirement for a high diversity of inputs rather than many inputs of the same type. This may render a node vulnerable if all inputs are essential, as well as prominent in terms of local trade relations. A high out-degree is still likely to imply that a node is influential in that it supplies many other sectors with materials. Hence, a change in the output of this node will impact on many others.

High betweenness centrality in this context is likely to indicate a high capacity to broker or control resource flows, but is also specifically relevant to connecting diverse parts of the economy. The application of these measures in our particular industrial context is discussed in more detail below.

### Centrality Results

Key processing industries, as measured by total degree, are the biological heavy intermediate compound producers, the non-biological heavy intermediate compound producers (inorganic chemicals), food processors, and steel and biodiesel production. These industries not only take in many local resources, and indeed are historically situated to make use of the region’s specific resources, but they also have local, as well as export, markets or routes for their numerous waste products. Arable agriculture, as the region’s original primary industry, again scores highly on total degree with many local markets in the food, biofuel and biochemical sectors. It scores highly on out-degree as we might expect, but it is also the means by which some local biological recycling loops are closed via animal feed manufacture, compost or digestate (as we would expect for such low-value, low-density products) increasing its betweenness centrality. The port and the nodes representing import and export of resources all have very high degree. Again this is unsurprising as this is a key defining feature of the region geographically and consequently in its economic structure. Waste aggregators are also shown to be important nodes by this measure. As generalists they take in numerous waste streams from many sectors and provide output to many sectors (scoring highly on in- and out-degrees), thus performing a key role in the closing of local resource loops. This is also reflected in their high betweenness centrality score.

Nodes/sectors with high in-degree include large regional “sinks” such as landfill and CO₂ emitted to the atmosphere, or means by which material leaves the system from numerous sectors such as the port and general export of resources. Others with a high diversity of inputs

<table>
<thead>
<tr>
<th>Table 5</th>
<th>Gini Coefficients for Interview and Augmented Network Frequency Distributions</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gini Coefficient</td>
<td>Interview Network</td>
</tr>
<tr>
<td>Total Degree</td>
<td>0.4725</td>
</tr>
<tr>
<td>In-Degree</td>
<td>0.5201</td>
</tr>
<tr>
<td>Out-Degree</td>
<td>0.5798</td>
</tr>
<tr>
<td>Betweenness Centrality</td>
<td>0.7761</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Table 6</th>
<th>Spearman’s Rank Correlation Coefficients for Interview Network Nodes in Interview and Augmented Networks</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rank Correlation Coefficient</td>
<td>Degree</td>
</tr>
<tr>
<td></td>
<td>−0.13833</td>
</tr>
</tbody>
</table>
include waste aggregators (as before), food processors, and cement production (as described in the motif section above). Interestingly, construction also scores highly. Again it has many local material connections due to the proximity of the port and the low value/weight ratio of its input streams, but it is also the recipient of many waste products such as hardcore from construction demolition, soil from composters or bottom fly ash from incinerators and power stations.

“Biological heavy intermediate compound” producers have retained the uniformly high centrality measures that they demonstrated from the interview network data alone, indicating that they are indeed key local industries. As visible in the network graph, companies in this sector take in biological raw materials and transform them to inputs for other manufacturing, buying in virgin biomass from processors, aggregators, agriculture, and importing. Many of them have their own dedicated heat and power plants visible in the circular power links to self. They produce bio-based waste that is sent to composting, cement kilns or exported to anaerobic digestion in other countries. Bio-based waste links to self are also visible where waste is recovered and reintegrated into the process on-site. Many companies of this type are very motivated to have “ground-to-ground” stories for their products and so the majority of their waste streams are not going to landfill (avoiding landfill tax also generates significant operational cost savings). These high scores may of course be due to the detailed interviews which we conducted within this sector, but it does also reflect the nature of biologically based industries and the wide diversity of material sources and product or waste destinations that are available to them. Material can be sourced from agriculture, bioprocessors, or imports, or as waste from food processing or other bio-based industries. Wastes can be used to generate energy or landfilled in conventional fashion but can also be composted, sent to anaerobic digestion, used to make biofuels, and ultimately can return to agriculture. Similarly, although they cannot take in waste products food processors have a high diversity of both input and output connections available to them. Both industries have a high betweenness centrality, indicating that they act as important bridges between different parts of the network. Other such bridges are the port, which connects many “imports” and “exports” to the regional network and other sectors through which resources are cycled such as agriculture, biodiesel production from waste, waste aggregators, and construction.

Agriculture seems to be retaining its important historic position within the area and is seen to be feeding many of the regional industries including a strong food processing sector. Food processing is linked to another characteristic regional industry, fuel and oil, via waste. In particular, oily food waste (an increasingly high-value product), which was previously refined to go into animal feed, is increasingly used to make biodiesel. The presence of petrochemical refining and specialist storage capacity means that it can be sent for local blending with diesel to go out into road fuel.

Other key industries in the region such as the oil refineries finally make an appearance in the analysis of this combined network; presumably as an important link between the biofuels industries and their many feedstocks and export from the system as final products. More nuanced analysis in the context of stakeholders’ concerns will be required to develop the usefulness of this approach further. It is interesting to note, however, that some industries which are considered extremely important to the region, such as oil refining and power production, appear only minimally or not at all by these measures of nodes’

<table>
<thead>
<tr>
<th>TABLE 7</th>
</tr>
</thead>
<tbody>
<tr>
<td>Top 10 Industrial Types in Augmented Network Ranked by Degree, In-Degree, Out-Degree, and Betweenness Centrality</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Degree</th>
<th>In-Degree</th>
<th>Out-Degree</th>
<th>Betweenness Centrality</th>
</tr>
</thead>
<tbody>
<tr>
<td>Biological heavy intermediate</td>
<td>Export: Resources</td>
<td>Biological heavy intermediate</td>
<td>Biological heavy intermediate</td>
</tr>
<tr>
<td>Waste Aggregator B: Large regional</td>
<td>Waste Aggregator B: Large regional</td>
<td>Import: Resources</td>
<td>Port</td>
</tr>
<tr>
<td>Food producer/retailer</td>
<td>Landfill: Nonhazardous</td>
<td>Non-Biological Heavy Intermediate</td>
<td>Waste Aggregator B: Large regional</td>
</tr>
<tr>
<td>Port</td>
<td>Sink: Gases (inc. CO₂)</td>
<td>Food Producer/Retailer</td>
<td>Food producer/retailer</td>
</tr>
<tr>
<td>Nonbiological heavy intermediate</td>
<td>Food producer/retailer</td>
<td>Waste Aggregator B: Large Regional</td>
<td>Biodiesel Production: Waste</td>
</tr>
<tr>
<td>Import: Resources</td>
<td>Construction and related</td>
<td>Port</td>
<td>Construction and related</td>
</tr>
<tr>
<td>Agriculture: Arable</td>
<td>Agriculture: Arable</td>
<td>Agriculture: Arable</td>
<td>Refinery: Crude oil</td>
</tr>
<tr>
<td>Steel Manufacture</td>
<td>Biological heavy intermediate</td>
<td>Steel manufacture</td>
<td>Agriculture: Arable</td>
</tr>
<tr>
<td>Biodiesel Production: Waste</td>
<td>Other manufacturing: Regional (SME)</td>
<td>Extractive: Limestone and lime</td>
<td>Other manufacturing: Regional (SME)</td>
</tr>
<tr>
<td>Export: Resources</td>
<td>Cement works</td>
<td>Timber/wood products</td>
<td>Airport</td>
</tr>
</tbody>
</table>

© 2014 Wiley Periodicals, Inc.
DOI 10.1002/cplx
importance. This is probably predominantly for two reasons: First, links are unweighted meaning that the importance of very large sectors with a low diversity of material inputs and outputs will be underestimated. Second, the nature of this metabolic network representation means that highly economically important industries whose connections exist mainly in the social and economic realm will not be ascribed their due importance. When additional layers of the network are added this picture may change.

**Verification**

Although we have indications from our preliminary analysis that our network augmented with industrial network archetype has a more realistic structure than interview network alone, further verification is clearly needed. Given the data restrictions that prompted the use of this methodology in the first place, it will not be possible to test our network against the real network directly. As alluded to above, however, this type of network representation, although incomplete, can be an important aid to promoting stakeholder engagement and can provide an object around which discussion of “whole system” issues can be based. Its compelling visual nature and the fact that it renders the usefulness of such models more directly apparent can be leveraged to encourage more connection data to be divulged. For example, the potential for the indirect impact on any given sector of removing a key node can be simply and interactively demonstrated, even though nodes and edges may be missing. Thus far, we have presented the model at a short session with key stakeholders at a workshop in the region. Initial feedback on the network approach and the structure of the network was positive. Following on from this, a series of workshop sessions as well as an interactive website are currently being developed with explicit reference to key network-related regional issues. Stakeholders, including those represented in the network and other regional experts without metabolic connections, will be invited to give feedback both on the forms of their own industrial archetypes and the properties of the network more generally. A full network analysis and development of models based on a verified network structure will follow this procedure.

**DISCUSSION AND FURTHER WORK**

One of the principal uses of a network structure such as this will be to explore potential dynamics on the network and how they relate to issues of interest to stakeholders in the area; in particular the connection of scenarios with qualitative regimes of network behavior such as high or low resilience or robustness and other system indicators meaningful to stakeholders. The dynamics of the existence and magnitude of links or flows in the system will be strongly influenced by the landscape of policy instruments in which they exist and different policy instruments may act differently on particular node types (e.g., European Waste Catalogue (EWC) codes or Renewable Obligation Certificates (ROCs)) or link types (e.g., carbon or landfill taxes). Recent and proposed changes in the regulatory environment, for example ROCs rebanding, or the change in form of the NISP, from a free to a subscription service, could hence have impacts which would percolate through the network differentially giving rise to unexpected consequences. Additionally, the interaction of several policy instruments on the network could have possible indirect effects which could also be explored. Policy effects on the network will combine with key price link dynamics, such as the links between virgin rape seed oil, waste food oils and virgin crude prices and the impact of exogenous shocks and stresses, such as increasing or highly variable fuel prices, or effects due to climate change and variability, such as unexpectedly high prices for wheat or other agricultural commodities with a strong impact on the bio-based economy. Wheat supply problems are already known to be an issue for the region’s bioethanol industry and variable fossil fuel prices have been identified as one of the key system drivers in the course of stakeholder workshops. Clearly, work is at early stages, but this network provides a realistically structured interaction space on which to test the impacts and interactions of proposed policy instruments and potential shocks and stresses on the system.

Further structural analysis could potentially allow us to identify which industry types are keystone nodes for the network and hence, from the context of regional decision making, what points of intervention can be identified. Work could also incorporate potential network reconfiguration by considering the impact of the addition or deletion of nodes and links under particular circumstances. This will allow us to represent upcoming changes in the region. The Humber area has a target for the production of 350 MW of renewable energy by 2021 and Yorkshire and Humber a regional target of 1,850 MW of renewable power. This will involve substantial network reconfiguration and addition of new industry types. Such changes and their impacts on the extant industrial network are clearly of interest to regional stakeholders.

Two large bioethanol plants are currently being commissioned and should be in full production by the end of 2013, with additional plants having been given planning permission. In total, more than 600 million liters of bioethanol per year, 50% of UK supply, will be produced in the region. Additionally Drax, the UK’s largest coal-fired and co-firing power station aims to switch 50% of its capacity to biomass only by 2017 with the enormous requirement for dry biomass supplies which will entail; offshore wind projects are expected to begin construction and...
significant development in technologies for biorefineries for the production of high-value chemicals from waste or other biological stream is on the horizon. Each of these scenarios will create the possibility of new types and magnitudes of connections between industries and the addition of new nodes or new types of node to the Humber network.

Further work to extract social/informational networks from our interviews is underway. This will allow interaction of different social network structures or network policy instruments (e.g., NISP) with other instruments of different types. All of our continuing work will form part of an on-going process with stakeholder feedback, participation and new data. The scenarios presented above, such as the effects of dynamic biomass prices within a nascent regional bio-based economy, and the identification of keystone companies within an industrial ecosystem, will be the immediate future focus of the presented research.

CONCLUSIONS

Tools from complexity science have the potential to aid decision makers managing real world complex socio-technical systems. Network visualization and analysis could be a helpful tool for such systems. However, in many real social or commercial situations complete network data are unavailable. This article has presented a novel methodology for constructing a regional industrial metabolic network based on coding of semistructured stakeholder interviews and incorporation of “network archetypes” for different industrial sectors. This approach allows us to collate and transfer incomplete, qualitative data into an easily visualized network form. The network visualization presents a readily accessible regional metabolic “portrait” which can be used in a participatory or one-to-one context to elicit further information from stakeholders and to discuss potential system dynamics.

Network data based on interviews from snowball sampling will suffer from inherent sampling bias which could give interviewed companies and their contacts higher degrees and centralities than they merit. Conventional methods of dealing with incomplete network data by simulating random graphs based on data from complete ego networks of a subsample are not suitable for our data or our modeling requirements. Industrial network archetypes overcome these limitations by using expert knowledge on characteristic connection structures. Including network archetypes reduces sampling bias but cannot completely remove it. Despite this, preliminary analysis of the resulting network allows us to identify key industrial sectors in the region based on their metabolic connectivity or centrality. Bio-based industries, which have a high diversity of inputs and outputs, were found to be highly central, along with the port, a crucial link in supply chains, and industries which are involved in cycling materials from waste product to resource such as agriculture, waste aggregators, biodiesel production, and construction. Other key regional industries such as oil refining were also found to be highly central. However, power production, an important characteristic of the area, did not score highly. This is likely to be due to its metabolic character of direct imports of fuel, single, simple process then export to grid, with minimal connections to neighboring industries. Further development and analysis of the network and its structure will be an on-going process with stakeholder feedback to ensure its relevance to the case study. Ultimately, we intend to use the network produced as an interaction structure on which to model the effects and interactions of different policy instruments and external shocks and stresses on the system as a whole and the companies within it. It is our contention that hybrid network construction from real data and generic archetypes could be a useful step to constructing network models of industrial (and other) ecosystems when data is incomplete, as well as providing visually compelling representations which both encourage and enhance further stakeholder engagement.

ACKNOWLEDGMENTS

The authors thank Dr. Jen Badham for useful discussion, the anonymous reviewers for helpful comments, and in particular the numerous participants from the Humber region who gave their time and shared their knowledge in the course of the work. The Evolution and Resilience of Industrial Ecosystems project at the University of Surrey is funded by the EPSRC “Complexity Science for the Real World” initiative.

References

1. Frosch, R.A.; Gallopoulos, N.E. Strategies for manufacturing. Sci Am 1989, 261, 144–152.
17. van Berkel, R. Regional resource synergies for sustainable development in heavy industrial areas: An overview of opportunities and experiences; Curtin University of Technology: Perth (Australia), 2006.
44. Markusen, A. Sticky places in slippery space: A typology of industrial districts. Econ Geogr 1996, 72, 293–313.
60. Freeman, L. A set of measures of centrality based upon betweenness. Sociometry 1977, 40, 35–41.