

STREAM: Social Networks

TITLE: Using Bayesian Agents to Enable Distributed Network Knowledge: A Critique

by ANET POTGIETER, KURT A. APRIL & RICHARD J. E. COOKE

Dr. Anet Potgieter

Senior Lecturer
Department of Computer Science
Computer Science Building, Main Campus
University of Cape Town
Private Bag
Rondebosch, 7701
Cape Town, South Africa

Telephone : +27 21 650-4056
Fax : +27 21 689-9465
Email : anet@cs.uct.ac.za

Dr. Kurt A. April

Professor in Leadership and Knowledge Management
Graduate School of Business
Breakwater Campus: Greenpoint
University of Cape Town
Private Bag
Rondebosch, 7701
Cape Town, South Africa

Telephone : +27 21 794-0261
Fax : +27 21 794-8938
Email : aprilkur@gsb.uct.ac.za

Richard J.E. Cooke

Senior Researcher
Department of Computer Science
Computer Science Building, Main Campus
University of Cape Town
Private Bag
Rondebosch, 7701
Cape Town, South Africa

Telephone : +27 21 650-5069
Fax : +27 21 689-9465
Email : rcooke@cs.uct.ac.za

Abstract

Resource based theory (RBT) states that there are dynamic relationships between individual-resource interactions, which ultimately determine an organisation's global behaviour in its environment. When combining in idiosyncratic, functional ways to enable an organisation's global behaviour, we call them complementary resource combinations (CRCs), and socially complex resource combinations (SRCs) when referring to only the complex web of social interactions of these resources. Casual ambiguity refers to the inherent uncertainty when the global behaviour is both tangibly evident and known, but the way in which the unique local interactions between SRCs amongst themselves and the environment ultimately contribute to the global behaviour is often unclear.

Thus, in order to understand social complexity and causal ambiguity of an organization, the SRCs emergent behaviours and the causal local interactions must be observed over time, and the inter-relationships must be identified and made tangible.

In our research, we use simple agents to observe the local and global behaviours, to mine the inter-relationships and to model the SRCs. These agents are organized into two types of agencies: Bayesian agencies and competence agencies.

The Bayesian agencies are the observers – they collectively implement specialised, distributed Bayesian networks, which enable the agencies to collectively mine relationships between emergent global behaviours and the local interactions that caused them to occur. The competence agencies are the actors – they use the beliefs of selected Bayesian agencies and perform dynamic network analysis. In dynamic network analysis, temporal data is used to predict changes that will occur in the SRCs. Most importantly, the Bayesian agencies observe and mine temporal patterns in various metrics over time, and the competence agencies evolve the SRCs.

Relationships discovered and maintained by Bayesian agencies and competence agencies are integrated into cutting-edge, resource-based topic maps (*ISO 13250:2002*), which provide a way of modelling the SRCs.

Keywords: Bayesian agencies; competence agencies; socially complex resource combinations; causal ambiguity; time-dependent emergent behaviour; topic maps

Introduction

Social networks are complex systems that are characterised by high numbers of interconnected component entities and a high degree of interaction between these entities. The interrelationships in such a network are dynamic and evolve over time. Temporal changes in social networks are difficult to understand and anticipate. The interrelationships between the component entities in a social network and its global behaviour can be so numerous and mostly hidden, and can affect so many different entities throughout the social network that it becomes extremely difficult to comprehend. Complexity theory is ideally suited to study social networks. Complex adaptive systems theory is a branch of complexity theory that studies systems that consist of agents that are collectively able to evolve in response to environmental changes. The agents in such a system constantly act and react to the actions of other agents and events in the environment. A social network is a complex adaptive system, in which people are agents interacting with each other. In this paper we describe how we use simple software agents to observe and model social networks over time. We based our research on Marvin Minsky's model of consciousness. Our agents observe local behaviours between people in social networks over time and global temporal characteristics (metrics) of these networks. The software agents model these interrelationships using Bayesian networks and topic maps, making hidden social relationships tangible.

Resource-Based Theory (RBT)

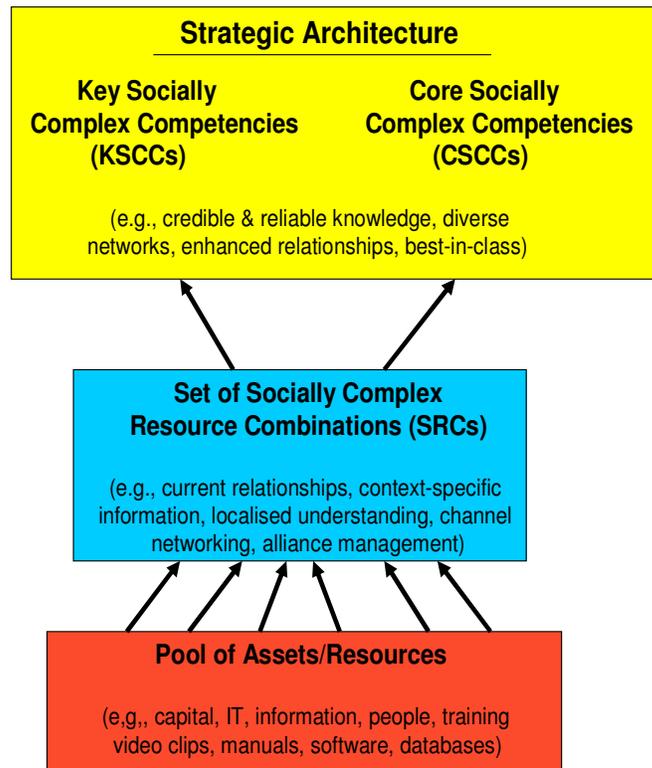
Our 21st century, post modernist, context raises important questions as to how value, for organisations and their customers, is created. Conventional management thinking and practice are being challenged at all levels, and postmodernism is redefining how we think about organisations, i.e., as complex adaptive enterprises, glued together through evolving, socially constructed and diverse knowledge and realities. It suggests reality is not out there to be discovered, but is rather embedded in the language, behaviours, and experiences of people in loose- and tight relationships, as well as our organisational relationships. An uncertain future implies that an organisation has no knowledge of either future problems or their possible solutions – it follows then that understanding, and the ability of an organisation to learn quickly, are the only appropriate methods of dealing with the uncertainty of future conditions. Understanding sources of dynamic learning, that lead to sustainability, have become major areas of research. With the emergence of the concepts of strategically architected key- and core capabilities (Potgieter, April and Bishop, 2005; April and Ahmadi-Izadi, 2004; April, 2002; April and Cradock, 2000), the focus of attention has shifted from outside the organisation to inside the organisation, i.e., the intra-organisation in its distributed- and networked sense. It has shifted from purely understanding the environmental and external factors to understanding the enablers of capability, systems and work processes, organisational structures and culture, as well as resource combinations.

Strategic Architecture

A department, or division, and its unique bundle of linked, idiosyncratic resources (Day, 1994), particularly its people/actor/agent resources, are the focal level of analysis in RBT. Organisational departments operate with substantially differentiated bundles of resources and assets (Amit & Schoemaker, 1993; Grant, 1991), using disparate approaches (April, 2002), and the individuals in them enact their differentiated identities through their skills, knowledge, behaviours, experience and relationships. Our underlying RBT-based framework (Figure 1) provides the basis for departments, and

individuals making up departments, to understand themselves better, perform their tasks more effectively, and be more responsive towards their clients and changing competitive environments.

Figure 1. Assets combine to make SRCs, that serve as bases for competitive advantage when organisations compete on competencies



An organisation's resources include all inputs, developed inside the organisation or acquired in the market, that allow it to work and to implement its strategy. We classify the organisational architecture into three layers:

- ◆ **Resources/Assets** are stocks of available factors that are owned or controlled by the organisation (Amit & Schoemaker, 1993; Dierickx & Cool, 1989). Assets, easily acquired on factor markets, can be tangible/visible (e.g., computer systems, software packages, infrastructure, capital, economic data and information, planning and accounting models, capital equipment, patents, articulated and codified knowledge) (April & Cradock, 2000; Schulze, 1994; Bogaert, Maertens & Van Cauwenbergh, 1994) or intangible/invisible (e.g., IT skills, organisational skills, team work-skills, educational skills). In reality, it is difficult to find situations in which competitive advantage is derived exclusively from a single strategic resource. Although these input factors are necessary to run a company, they, solely, provide no lasting competitive advantage (necessary, but not sufficient to take leading marketplace positions).

- ◆ **Complementary Resource Combinations (CRCs)** results from bundles, or combinations, of certain assets (April and Cradock, 2000), which exhibits enhanced networked-complementarity in deployment or application (Barnard 1938). CRCs are complex combinations, not factor inputs, of resources – people, technology, and business processes, with specific and sophisticated networks of interrelationships, that organisations use to transform factor inputs to value-generating outputs. Finely-honed CRCs, a source of competitive advantage, enable an organisation to take the same factor inputs as competitors and convert them into products and services, either with greater efficiency and flexibility in the process or with greater quality in the output. It is only when resources are transformed, combined or applied, that these generic resources become part of an organisation's CRCs, contributing directly to the development of distinctive competences. Social complexity refers to the complex behaviour exhibited by individuals in a complex adaptive enterprise, when its CRCs are embedded in a complex web of social interactions. These CRCs are referred to as socially complex resource combinations (SRCs) (Potgieter, April & Bishop, 2005; April, 2002). The organisation's department- or division-specific characteristics - specific routines and procedures, specific information, and current relationships between the people, the internal culture, etc., make the SRCs unique to that department and they cannot simply be recreated by another department or division.
- ◆ **Competencies** are value-generating capabilities of an organisation, which enables it to compete both now, and in the future. Socially complex competencies (SCCs) are emergent sets of coordinated networks, that efficiently and effectively leverages distributed, inter- and intra-organisational processes, individual skill-sets, accumulated knowledge, and coordinated, patterned behaviour, which ultimately enables an organisation to be a “player” in the “market-game”. Those SCCs which are merely necessary for the department, or division, to be a player in their field of activity in the marketplace, are termed *key socially complex competencies (KSCCs)*. Those SCCs that set the department or division apart from other departments or divisions, and therefore justify the existence of that department or division, are termed *core socially complex competencies (CSCCs)*.

The main distinction between assets and competencies is that assets are related to *having, and are diminished by their use and valency* while competencies are related to *doing and understanding, and are enhanced by their continued, appropriate use*. The more a competency is utilised, the more it can be refined and the more sophisticated and difficult to imitate it becomes. This characteristic manifests the dynamic or evolving characteristic of competencies, and therefore makes it difficult to understand with static or non-distributed models. Up until recently, systems like computers or simulators were unable to recreate what is essentially the “organisational, collective memory of doing”.

Sustainability

Most discussions of sustainability focus on defensive strategies based on existing resource strengths, so a lot of time is spent on understanding differing histories of strategic choice and performance, because management of these departments or divisions appear to seek asymmetric positions, and because of the various routines it has developed to manage them (Teece, Pisano and Shuen, 1990). Existing resource strengths are obviously important, but securing the long-term future of an organisation must consider how to derive unique complementary areas of value added for the future. Both the concepts of sustainability and RBT, as posited in the current academic literature though, can limit organisations in understanding the full nature and dynamics of

strategy. Firstly, sustainability is a journey and not a destination and it only becomes meaningful when the journey is experienced. Given the changing local and global environments, the process of identifying the journey, through adaptive awareness, presents the main challenge. Secondly, RBT (as was currently known up to now) attempts to identify and nurture the resources that lead to sustainability – but dynamic environments constantly call for a new generation of CRCs that are heterogeneously distributed throughout the organisation, and the identification and ultimate grafting out of redundant, obsolete and valueless CRCs. Hence, it is our thesis that previous work in the RBT area has focused too little attention on the richness and diversity of socially complex resource combinations, who through their interactions bring to bear on the value-adding organisational activities, constantly renegotiated flows of knowledge, learning, cumulative experience, capability accumulation and integration of functional capabilities. Sustainability is a dynamic process, dependent on people, rather than a set of activities locked in time.

Social Complexity

Organisations are not atomistic agents, instead they are recurring and dynamic agent linkages, and embedded in networks that influence competitive actions (Granovetter, 1985, 1992; Burt, 1992). The “structure of any social organisation can be thought of as a network” (Nohria & Eccles, 1992: 288; Lincoln, 1982) that operates and is operated on in an environment which itself is an environment of other organisations (Van Wijk, Van Den Bosch & Volberda, 2003), and the actions of individuals within the network (network actors/agents) are shaped and constrained because of their position and embeddedness in the network (Nohria, 1992). With the emergence of knowledge as a strategic asset (Grant, 1996; Winter, 1987), knowledge transfer in value-generating ties have become a focal point for research, which in themselves may be (a) assets, (b) information (access, timing or referrals), and (c) status (Van Wijk, Van Den Bosch & Volberda, 2003; Burt, 1992; Galaskiewicz, 1979). According to Van Wijk, Van Den Bosch & Volberda (2003: 430) ‘access’ means that network ties influence access to parties; ‘timing’ allows agents to obtain information sooner than it becomes available to agents without such contacts; and ‘referrals’ constitute processes which provide information to agents in the network on available opportunities. Hence, the information benefits of ties influence network formation, and consequently, opportunities to combine and exchange knowledge (Van Wijk, Van Den Bosch & Volberda, 2003; Dyer & Singh, 1998; Nahapiet & Ghoshal, 1998). In addition, the organisational mode – singular, co-located team-based, virtual team-based, networked intra-organisationally within-industry, networked intra-organisationally trans-industry, networked trans-organisationally within-industry, networked trans-organisationally trans-industry – through which individuals cooperate, affects the knowledge they apply, and are able to apply, to business activity. These modes affect both: (a) the way in which *static* (i.e., presently possessed) knowledge is employed (‘knowledge execution and substitution effects’), and (b) the *dynamics* of future knowledge acquisition and response to new developments (‘knowledge flexibility effect’) (Conner & Prahalad, 2002). In effect, the individual chooses a knowledge cost-path, for current and future success and viability (measured as more valuable understanding to business activity), between what is available in the company, what may become available in the company, what is available outside of the company, and what may become available outside of the company, and blending knowledge substitution effects with knowledge flexibility effects, i.e., an individual’s “true colleagues” and viable network could exist inside or outside, or inside and outside of the formal structures of the organisation. Loasby (1999: 58) claims that the development of knowledge is a path-

dependent process, in which “the acquisition of certain kinds of ‘knowledge how’ facilitates the acquisition of further knowledge of the same kind, and impedes the acquisition of knowledge of incompatible kinds; and this principle applies both to the performance of productive operations and to the procedures by which we seek to develop new ‘knowledge that’.” Each person’s cognitive repertoire develops within, and across, institutional contexts, and become to a greater and lesser extent dependent on those contexts, both for structure, applicability, permeability and ultimately flexibility. Penrose (1959, 1995: 53) suggests that individuals are likely to develop capabilities with very different breadths of application, some for example being useful only within a particular firm, some within a group of firms, some within particular industries (e.g., manufacturing), and some in any productive activity. In social complexity, the source of competitive advantage is known, but the method of replicating the advantage is unclear and casually ambiguous, because of the huge number of choice-variables that exists in the correlation of initial knowledge endowments, individual choices, mitigating company moods, contexts, and organisational modes. Examples of mitigating company moods include corporate culture, the interpersonal relations among managers or employees in an organisation and trust between management and employees. Competencies and ‘knowing how’ is not a single-track disposition like a reflex or habit, rather its exercises are observances of rules or canons or the application of criteria (Conner & Prahalad, 2002; Ryle, 1949). Moreover, competencies may be, and clusters of competencies always are, clusters of dispositions, which may be realised in many ways; and, through the exercise of intelligence, they may be developed in ways which are not foreseeable (Conner & Prahalad, 2002: 519). The limitations on their development cannot be foreseen either, for, as Nelson & Winter (1982: 84) point out, “performance takes place in a context set by the values of a large number of variables; the effectiveness of the performance depends on those variables being in appropriate ranges” – we thus cannot fully know the ambiguity of scope presented by all the possible ranges available, particularly as the relevant variables are (a) not known to us, (b) new variables are not yet known to us, and (c) the importance and quality of the variables at any given time is inherently dynamic. Socially complex resource combinations (SRCs) therefore depend upon large numbers of networked-people or networked-teams engaged in co-ordinated, emergent action such that few individuals, if any, have sufficient breadth of knowledge to grasp the overall phenomenon. Thus, in order to understand social complexity and causal ambiguity of an organisation, the SRCs’ emergent behaviours, the correlated variables and the causal local interactions must be observed over time and their interrelationships must be identified and made tangible.

Social Networks

A social network is a representation of people and their relationships at a point in time as a mathematical graph. The graph’s set of vertices represents people and its set of edges represents relations between people. Social network analysis (SNA) is the study of social networks, founded by sociologists. The individual actors in an organisation’s social network are linked in a complex web of relationships that change over time. These relationships between actors emerge, strengthen and decay as a result of individuals’ positions in the network, their behaviour and the influence of the organisational environment. The capability to predict these changes in relationships before they occur is highly beneficial to an organisation. Examples of the advantages of such social clairvoyance include:

- Being able to acquire, discard or reallocate resources to meet the new requirements of actors in the changing organisational structure.
- Predicting potential problems (e.g. losing a key actor, social bottlenecks) in the future network and acting to eliminate them before they occur.
- Planning an elegant communication infrastructure to handle the increased or decreased demands of communication links.
- Identifying the structure of a criminal network (i.e. predicting missing links in a criminal network using incomplete data).

Temporal considerations in social networks

Predicting changes to a social network is called “the link prediction problem”. Liben-Nowell and Kleinberg (2003) explain it as:

Given a snapshot of a social network at time t , we seek to accurately predict the edges that will be added to the network during the interval from time t to a given future time t' .

This approach to the problem is limited, however, as it attempts to predict the evolution of a complex entity over time from a snapshot. Consider the analogy of trying to predict the position of a thrown ball a second from now, given only a photograph of the ball when it was released from the thrower’s hand. It is true that the ball’s position can be approximately predicted, but it would be better to have seen the ball move through the air. In other words, we need to know the velocity of a social network, not just its position. To continue the analogy, the position of a social network is given by traditional SNA metrics calculated from a snapshot, but velocity can be determined only by calculating temporal DNA metrics using the a history of changes to a network (i.e. an “animation” of the network over discrete time intervals).

In our research we define useful temporal metrics by extending the definition of static metrics to account for time. We use agents in a complex adaptive system to observe and learn from these metrics over time. We are also interested to investigate whether temporal metrics can be used to determine where links are missing in incomplete data. For instance, given partial intelligence data describing a criminal network, can we determine which criminals know each other, even though the link is not given as a datum? This problem is a simple alteration of the link prediction problem. The alteration is that instead of trying to predict future links based on all the networks up to the present, we rather are trying to predict all the present links (including potential missing links) based on all networks up to, but not including, the present network.

Review of SNA and DNA techniques

Although SNA has existed for over fifty years, most analysis techniques have been designed for static data. For example, the primary reference for social networks is Wassermann and Faust’s *Social Network Analysis* (Wassermann and Faust, 1994), cited by the majority of social networks research papers. This book contains no mention of temporal metrics, even though it was written in 1994 when electronic networks were well established. It is difficult to collect social data for numerous individuals by hand using survey techniques. However, with the increase in the use of computers, collecting enough data to create numerous graphs over fixed time intervals becomes possible. An example is creating a graph per week from email data, using a server’s email log of “to”, “from”, and “date” fields (Campbell, Maglio and Cozzi, 2003). This series of graphs can be used to study the evolution of the network and the change over time in various metrics.

The analysis of the changes in social networks over time is called dynamic network analysis (DNA). It is currently a popular avenue of research for law enforcement and intelligence agencies, given the rise in the global activities of terrorists and other organised criminal groups (Coffman, Greenblatt and Marcus, 2004). Such groups have been labelled “dark networks”, and their structure and behaviour differs widely from normal social networks. For example, they trade efficiency for secrecy in structure and have unusual patterns of communication (Fellman and Wright, 2004). Carley is one of the most prolific researchers in the modelling of dark networks using dynamic techniques. She has created a dynamic network program, *DyNet*, where multiple agents model the social behaviour of human beings, with access to resources and organisations (Carley, forthcoming). This program is used to understand network evolution and the best way to destabilise terrorist networks. These techniques are powerful, but relatively domain specific and complex. There have also been a few purely theoretical studies done on the change of the structure of networks over time. Holme’s work has focussed on this, including studies on the changing metrics of an Internet dating network (Holme, Eding and Liljeros, 2004; Holme, 2003).

Existing metric and link prediction

This section provides a summary of traditional SNA metrics and prediction techniques. The following section explains how traditional metrics can be extended to take time into account.

Every traditional SNA metric is described and defined in *Social Network Analysis* and summarised in an online book by Hanneman (2001). Most metrics can be grouped into one of the following three categories:

- Power metrics (or centrality of a node): centrality; closeness; betweenness; indegree; outdegree; degree; number of messages sent/received; Bonacich’s power and centrality index.
- Aggregate graph metrics: size, complexity, density; number of paths of length 1, 2, 3, etc; connectedness; extent of triadic transitivity; diameter; mean flow; mean geodesic distance; Hubbel and Katz flow; Taylor measure; extent of reciprocity of attraction.
- Structural metrics: number and location of cliques, n-cliques, k-plexes, k-cores, components, blocks and cutvertices.

Existing link prediction techniques can use the values of these metrics in a graph instance to determine where new links are likely to arise. (For instance, it is more likely that a new link will be incident to a node with a high degree than a node with a low degree). However, there have been very few investigations of link prediction. Taskar et al. (2004) used relational Markov models to learn patterns of cliques and transitivity in webpages and hyperlinks. Popescul and Ungar (2003) made citation prediction systems using statistical learning that extended inductive logic programming. Their system learnt link prediction patterns from queries to a relational database, including joins, selections and aggregations. Both these prediction systems included node attributes (e.g. webpage text) in addition to relational features. This makes them more powerful than prediction systems using only SNA metrics, but also more domain specific. Liben-Nowell and Kleinberg (2003) tested the predictive power of only proximity metrics, including common neighbours, the Katz measure and variants of PageRank. They found some of these measures had a predictive accuracy of up to 50% (compared to a random prediction’s accuracy of less than a percent).

Proposed Temporal Metrics

Our proposed temporal metrics extend the SNA metrics listed in the previous section by combining them with simple statistics and financial statistics. The following paragraph discusses an example of temporally extending a metric – a technique that can be applied to any static metric.

Let us consider the outdegree of a node as an example metric. The outdegree of a node can be said to be a measurement of a person's influence over others. The static outdegree metric that is usually used in SNA shows a person's influence only at one point in time (i.e. in just one graph). To gain a more comprehensive understanding of the concept of outdegree we should consider also the change in outdegree over time. This allows us to see whether a person is becoming more or less influential. We first consider two metrics that determine a node's normal social influence: *outdegree-mean* and *outdegree-median*.

The *outdegree-mean* $\bar{O}_{m,n}(v_i)$ of a node, v_i , between time steps m and n , is the mean number of links from v_i to any node, calculated as:

$$\bar{O}_{m,n}(v_i) = (n-m)^{-1} \cdot \sum_{t=m}^n \#\{e_{i,j} : e_{i,j} \in E_t, v_i \in V_t\},$$

where E_t is the set of edges in the network at time t , and V_t is the set of vertices, and $\#$ denotes the number of links.

Outdegree-median can be calculated as:

$$\Theta_{m,n}(v_i) = \varphi_{\frac{n-m+1}{2}}, \text{ an element of } \theta, \text{ the ascending ordered set of the elements of}$$

$O_{m,n}(v_i)$, the outdegree values of v_i from time step m to time step n .

We can also calculate the spread of the outdegree values using *outdegree-variance*,

$$\text{defined as: } \underline{O}_{m,n}(v_i) = (n-m)^{-1} \cdot \sum_{t=m}^n (\#\{e_{i,j} : e_{i,j} \in E_t, v_i \in V_t\} - \bar{O}_{m,n}(v_i))^2.$$

This additional metric shows how variable a person's influence is over time (i.e. whether they are a stable socialite or if they oscillate between being highly extroverted and highly introverted with a very large *outdegree-variance*). Finally, we can calculate the direction of the change in time, which is perhaps the most important metric to consider if we are trying to predict whether this node will gain more links. For this, we borrow two measures of share return from finance (Ross et al., 2001).

Outdegree-return is defined as:

$$\text{outdegree}_m(v_i)^{-1} \cdot (\text{outdegree}_n(v_i) - \text{outdegree}_m(v_i)),$$

and is the percentage increase or decrease in outdegree over the entire period. This allows us to see whether a person is becoming more or less influential over time.

The second measure is *outdegree-return-average*, which is defined as:

$$\text{outdegree}_m(v_i)^{-1} \cdot \sum_{t=m}^n (\text{outdegree}_{t+1}(v_i) - \text{outdegree}_t(v_i)),$$

and shows the average percentage increase or decrease in outdegree per time step, over the entire period. These two metrics best separate the social shooting stars from the socially unambitious.

By applying the statistical techniques of mean, variance and return to static SNA metrics we obtain temporal metrics. We can create temporal metrics from both the category of static power metrics and the categories of aggregate graph metrics and structural metrics. The former category gives us a number of variables per node and the latter categories give us variables for the graph as a whole. Having both types of

variables (per node and per graph) allows a system to predict links not only by observing the changing relations between nodes, but also by observing the structural changes to the graph as a whole.

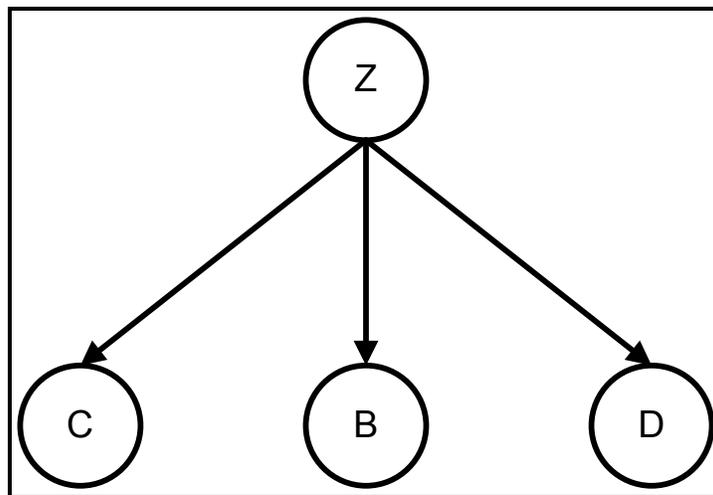
Modelling Uncertainty in Social Networks using Bayesian Networks

Bayesian Networks provide the ideal technology to reason about SRCs. These techniques are suitable for link prediction and have been used for a similar purpose (McNee et al., 2002; Taskar, Abbeel, Wong and Koller, 2004; Popescul and Ungar, 2003).

A Bayesian network is a directed acyclic graph (DAG) that consists of a set of nodes that are linked together by directional links. Each node represents a random variable or uncertain quantity. Each variable has a finite set of mutually exclusive propositions, called *states*. The links represent informational or causal dependencies among the variables, where a parent node is the *cause* and a child node the *effect*. The dependencies are given in terms of conditional probabilities of states that a node can have given the values of the parent nodes (Pearl, 1988). Each node has a conditional probability matrix to store these conditional probabilities, accumulated over time.

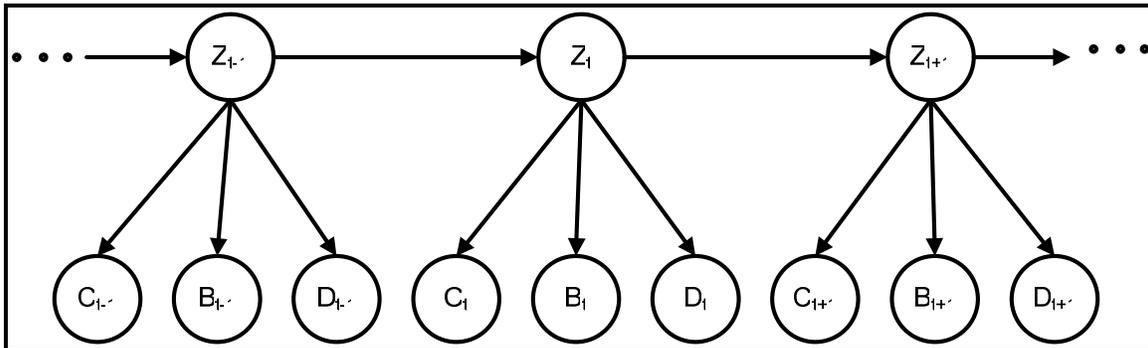
Figure 2 illustrates a simple Bayesian network that models the relationship between three SNA metrics, namely closeness (*C*), betweenness (*B*), degree (*D*) and one hidden variable, namely the class variable (*Z*). In Figure 2 below, the states of the hidden class variable *Z* are mined from historical data (calculations of *C*, *B* and *D*). The class variable *Z* is the single cause influencing multiple effects (*C*, *B* and *D*). This probability distribution is called a *naïve Bayes model* or sometimes called a *Bayesian classifier* (Russell & Norvig, 2003).

Figure 2. A Simple Static Bayesian Network



In a DNA, some metrics change over time. In dynamic Bayesian networks, multiple copies of the variables are represented, one for each time step (Pearl & Russell, 2000). A Bayesian network that does not model temporal relationships, is called a static Bayesian network. Figure 2 above is an example of a static Bayesian network. Figure 3 illustrates a dynamic Bayesian network that models the closeness, betweenness and degree SNA metrics at different time-steps.

Figure 3. A Dynamic Bayesian Network



Bayesian learning can be described as “mining” the structure of the network and calculating the conditional probability matrices from history data. The data may be incomplete and the structure of the Bayesian network can be unknown.

Bayesian inference is the process of calculating the posterior probability of a hypothesis H (involving a set of query variables) given some observed event (assignments of values to a set of evidence variables e),

$$P(H | e) = \frac{P(e | H)P(H)}{P(e)}$$

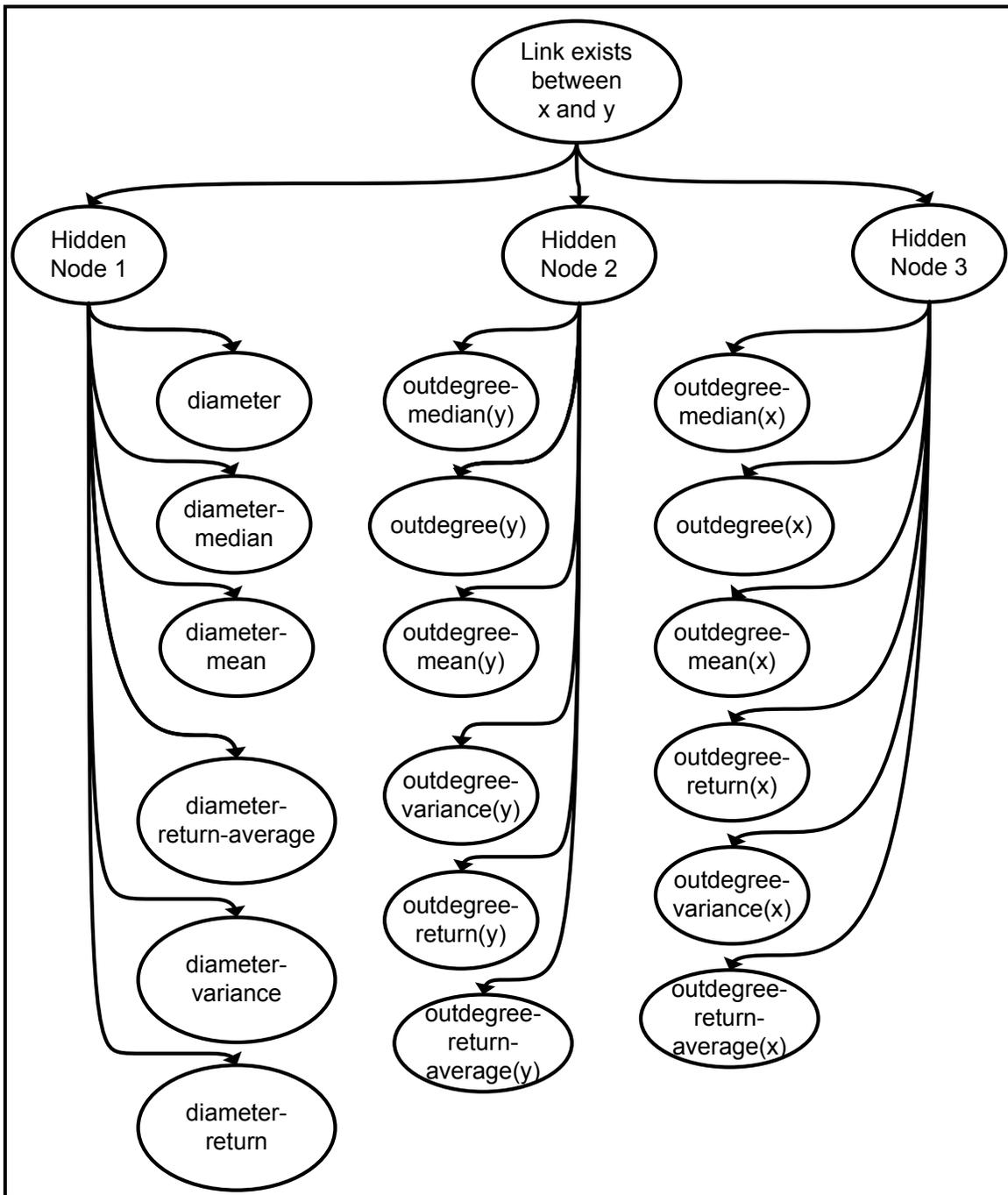
Both static and dynamic Bayesian networks can be used for link prediction in social networks.

Link Prediction using a Static Bayesian Network

A static Bayesian network can be used for link prediction by modelling cause-effect relationships between temporal metrics and the emergence of links between nodes in a social network. An example of such a Bayesian network is shown in Figure 4 below. The aim of this network is to learn what metrics are the best predictors of links occurring in the future, or that a link is missing if given incomplete data. In this simple example we can see that the target node at the top of the diagram represents the adjacency between two nodes, x and y . This is an observed variable. The three hidden nodes below it summarise (in this hypothetical case) the probabilities relating to the observed variables holding the values of the metrics for: the graph as a whole (hidden node 1); node y (hidden node 2); and node x (hidden node 3). In the implementation of our system, there will be over 150 various metrics relating to the graph and its nodes. Thus we could expect to see the Bayesian network shown below containing up to three hundred nodes.

We are hopeful that the large number of observed variables (five extensions to each of thirty static metrics) in a Bayesian network will create a highly accurate link prediction system.

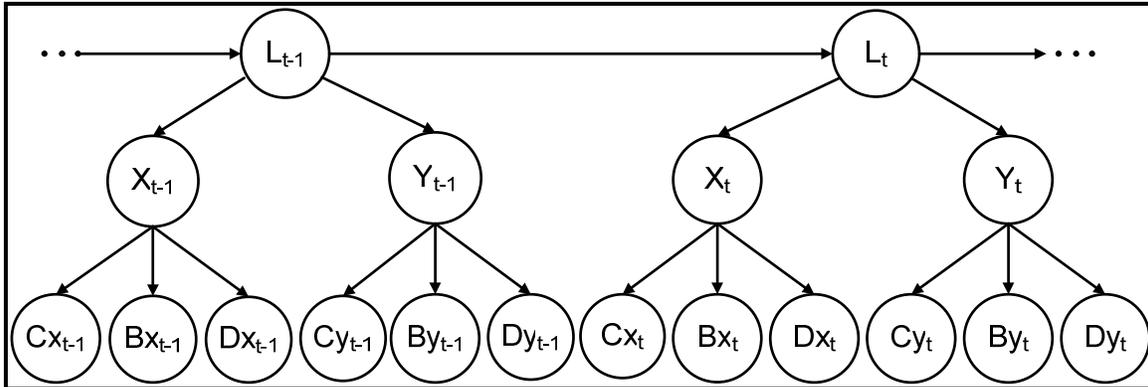
Figure 4. A Static Link Prediction Bayesian network



Link Prediction Using Dynamic Bayesian Network

A dynamic Bayesian network can be used for link prediction by modelling cause-effect relationships between SNA metrics and the emergence of links between nodes in a social network at different time-steps. An example of such a Bayesian network is shown in Figure 5 below. In this simple example we can see that the nodes at the top of the diagram ($\dots, L_{t-1}, L_t, \dots$) represents the adjacency between two nodes, x and y , in different time-steps ($\dots, t-1, t, \dots$). This is an observed variable. The two hidden nodes (X and Y) below each of these variables are naïve classifiers for the SNA metrics for that node.

Figure 5. A Dynamic Link Prediction Bayesian network



Agents in DNA

Though the use of multi-agent systems in DNA is relatively new, they have been used for many years in SNA. A multi-agent system normally uses multiple complex agents to complete a complicated task and can prove to be a highly simple and powerful solution to problems. Examples of the use of agents in SNA include determining trust and reputation (Pujol, Sanguesa, Delgado, 2002) and developing recommendation systems (Kautz, Selman and Shah, 1997). As agents are software programs, they can easily be extended to accommodate temporal network changes.

In our research we use simple agents, acting as part of a complex adaptive system, to observe and mine temporal relationships from social networks.

Agents and Agencies

Agents are commonly viewed as the next-generation model for engineering complex, distributed systems. There is, however, no consensus in the research community on what an agent is. Some researchers refer to single beings as agents (complex agents), while other researchers refer to independent components within a single being or system as agents (simple agents). Current research in multi-agent systems uses complex agents.

Minsky (1988) first established the concept of simple unintelligent agents combined into intelligent agencies. He describes the mind as a “society” of tiny components that are themselves mindless. He refers to each of these components as agents. His simple agents combine into (sub)societies, called agencies. The agencies are intelligent through the interaction amongst the (unintelligent) agents. These societies function as complex adaptive systems. Our research falls into this category, and differs from mainstream multi-agent systems.

According to Minsky (1988), an agent is:

Any part or process of the mind that by itself is simple enough to understand - even though the interactions among groups of such agents may produce phenomena that are much harder to understand

Simple and complex agents share one common concept, namely the concept of *agency*. Working with the concept of agencies rather than with the concept of agents, can help to create order out of the terminology chaos.

Agents are grouped into agencies, where the definition of an *agency* is:

any collection of simple agents considered in terms of what it can accomplish as a unit, without regard to what each of its constituent agents does by itself (adapted from Minsky, 1988).

An agent that can accomplish all its goals independently from other agents forms a single agency consisting of this particular agent as its sole member, having the same functionality as the agent.

Agents that collectively accomplish goals are grouped into agencies according to the functionality that they collectively achieve. If the collective global behaviour of the agencies is not obvious from the local behaviours of the agents, the agencies comprise a complex adaptive system.

In our research, we use the BaBe adaptive agent architecture (Potgieter, 2004; Potgieter, April and Bishop, 2005) consisting of simple agents that function as a complex adaptive system. Our agents observe the local and global behaviours, mine the inter-relationships and model the CRCs. These agents are organized into two types of agencies: Bayesian agencies and competence agencies. Our Bayesian agencies and competence agencies function as a complex adaptive system – they learn and adapt from what they observe.

The Bayesian agencies are the observers – they collectively mine relationships between emergent global behaviours and the local interactions that caused them to occur. The competence agencies are the actors – they use the beliefs of selected Bayesian agencies and perform dynamic network analysis. In dynamic network analysis, temporal data is used to predict changes that will occur in the CRCs. Most importantly, the Bayesian agencies observe and mine temporal patterns in various metrics over time, and the competence agencies evolve the CRCs. We have applied this architecture to model CRCs in retail environments, and we are currently researching the use of BaBe to mine and understand SRCs.

Achieving Self-Awareness

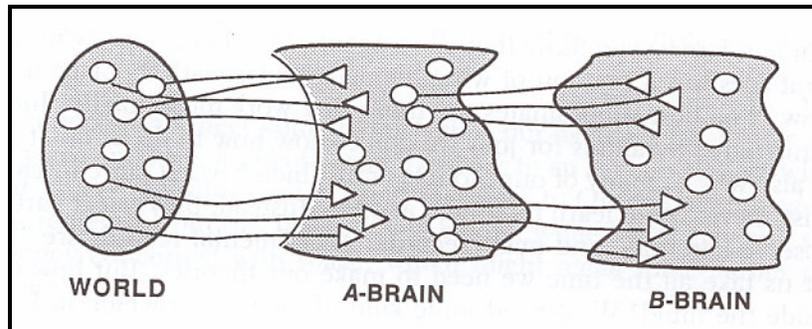
In order to achieve self-awareness in a social network, the local behaviours of the participants must be understood, and how they act together and interact with the environment to form the whole.

According to Baas and Emmeche [1997], understanding is related to the notion of explanation. A complex adaptive system uses the hyperstructures in its internal model for explanation and understanding. It uses observation mechanisms to create and maintain these hyperstructures. The process of adaptation relies heavily on the observation mechanisms and involves a progressive modification of the hyperstructures [Holland 1995].

The human mind is self-aware and capable of self-observation and self-interaction. Consciousness may be seen as an internal model maintained by the mind. In Marvin

Minsky's Society of Mind, he describes a model of consciousness [Minsky 1988]. In this model, observation mechanisms called *A-Brains* and *B-Brains* maintain internal models consisting of hyperstructures called K-Lines. Each K-Line is a wire-like structure that attaches itself to whichever mental agents are active when a problem is solved or a good idea is formed [Minsky 1988]. Minsky describes how a system can watch itself, using a *B-Brain*. In Figure 6, the A-Brain has inputs and outputs that are connected to the real world, and the B-Brain is connected to the A-Brain. The A-Brain can sense and influence what is happening in the world, and the B-Brain can see and influence what is happening inside the A-Brain.

Figure 6. Minsky's A-Brain and B-Brain [Minsky 1988]



Our Bayesian agencies form the A-Brain and they use dynamic Bayesian networks as hyperstructures. The competence agencies form the B-Brain and they use resource-based topic maps (*ISO 13250:2002*) as hyperstructures.

The A-Brain: Bayesian Agencies

Our Bayesian agencies consist of simple re-usable components, where each component can be one of three re-usable components, namely node components, link components and belief propagation agents. Collectively these simple components capture the knowledge in the social network by collectively implementing distributed Bayesian networks. Each node component implements a Bayesian network node. Each network link is implemented by a queue, together with a link component that participates in the synchronization of messages flowing to the child, or to the parent node via the queue.

For each queue, a belief propagation agent is deployed that listens on that queue for messages from the child or parent node of the associated network link. The Bayesian agents collectively perform Bayesian inference by using localised message passing in response to the environmental evidence in order to update beliefs of network nodes.

The Bayesian agencies implement the A-Brain that is connected to the real world. As soon as evidence is received from the environment, the Bayesian agents collectively perform Bayesian inference by using local message passing. Node components incrementally learn from evidence received from disparate data sources within the organisation or from external data sources.

The B-Brain: Competence Agencies

The competence agencies use the beliefs of selected Bayesian network nodes to determine if certain business components must be activated or not. Business components are re-usable components containing parts of business processes or workflow processes. Each competence agency monitors a set of constraints on the beliefs of a set of Bayesian network nodes – the constraint set. If all the constraints in a

constraint set are met, the competence agency can activate its associated business component. This business component can execute part of a business process or part of the workflow of the organisation.

The competence agencies can be viewed as constituting the “B-Brain”. These agencies can “see” inside the “A-Brain” by inspecting the beliefs of nodes and acting upon these beliefs and possibly changing the state of the environment, influencing the collective Bayesian inference of the Bayesian agencies – the “A-Brain”.

The competence agencies use topic maps to represent all knowledge regarding the CRCs. Topic maps provide a powerful formalism to represent meta-data. It is the new ISO standard (*ISO/IEC 13250:2002*). According to this standard, a topic map is:

A multidimensional topic space — a space in which the locations are topics, and in which the distances between topics are measurable in terms of the number of intervening topics which must be visited in order to get from one topic to another, and the kinds of relationships that define the path from one topic to another, if any, through the intervening topics, if any.

In the CRCs, the topics are resources in an organisation that are related to each other by associations. Occurrences relate resources in the organisation to information resources that contain information about them such as websites, documents, and rows in database tables, video clips, and so forth. These information resources are identified by URI's.

Topics, associations and occurrences can be organised into topic classes, association classes and occurrence classes. In SRCs, topics will be persons and topic classes will typically capture organisational structures, associations will identify the type of interactions between persons, and occurrence classes will relate to different kinds of information resources.

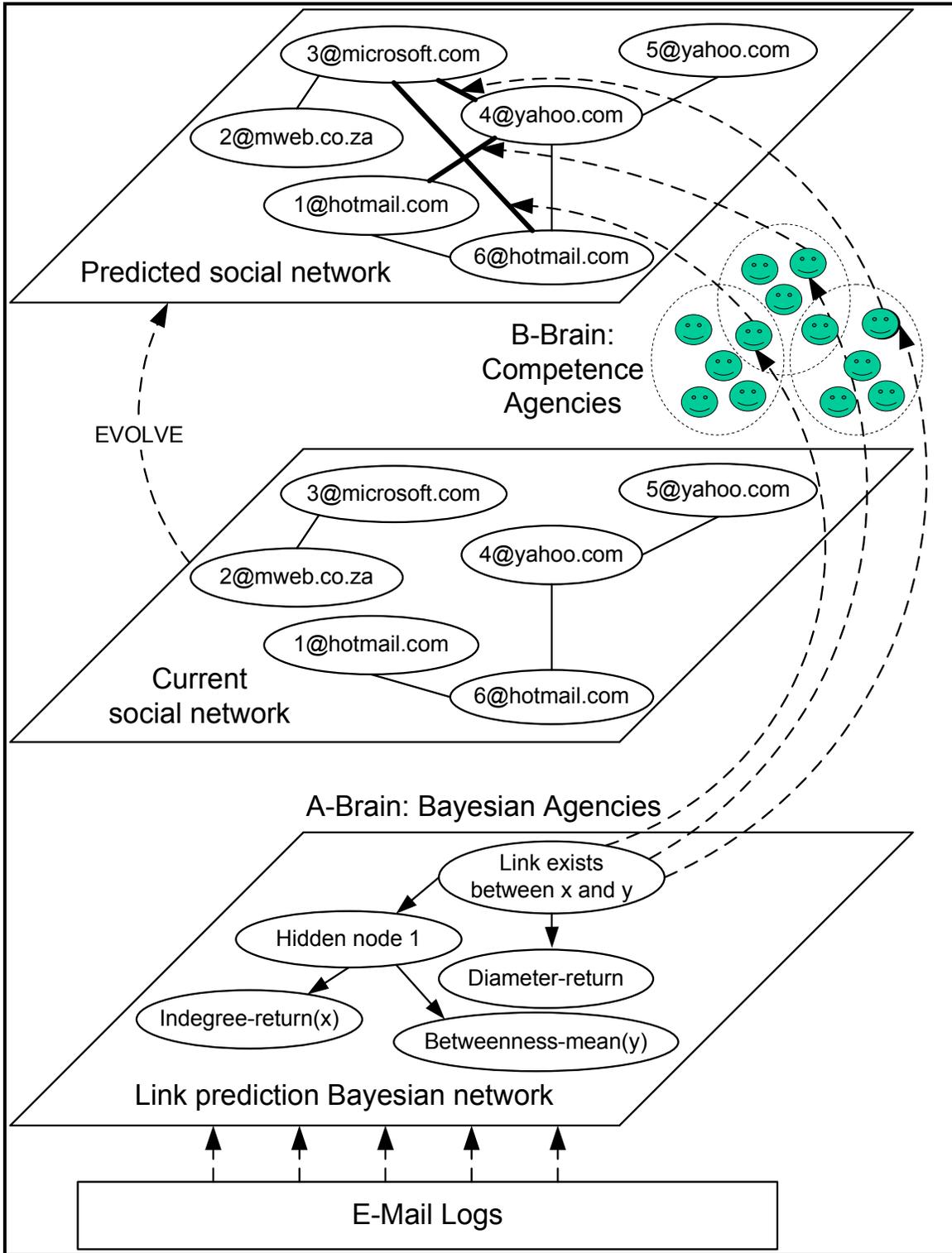
Link Prediction through Self-awareness

The competence agencies will strive to make implicit or hidden relationships explicit by querying the Bayesian Agencies and “evolving” the topic maps on an ongoing basis by integrating predicted social network links into the topic maps.

Figure 7 below illustrates how the Bayesian agencies and the competence agencies can achieve self-awareness as in Marvin Minsky's model of consciousness. In this example, the Bayesian agencies collectively implement a static Bayesian network using temporal metrics. These agents observe and mine patterns from the e-mail logs, while the competence agencies observe and query the beliefs of the Bayesian agents in order to discover emerging links in the social network over time. The competence agents then integrate these links into the social network represented using topic maps.

In this diagram, a static Bayesian network models the cause-effect relationships in temporal metrics. (It is of course only an example, and contains only a tiny fraction of the number of nodes that would be found in the real system). The first tier, at the bottom of the diagram, represents the raw social network data held in an email log. These data are updated over time as new emails are sent and are used to generate social networks in the tier above. For every time step (e.g. one per day, per week or per month), the social network will be evolved – new links will be added to the social network.

Figure 7: Proposed Link Prediction System



Future Work

In our future work, we want to test the system with Bayesian agencies implementing both the static Bayesian network with temporal metrics and the dynamic Bayesian network with static metrics, and benchmarking the performance of the Bayesian agents with the different models.

Although not a focus of our research, creating universal benchmarks for temporal metrics would be a fascinating area of future work. A temporal metric benchmark is a fixed value to which the metrics of various social networks in different stages of development can be compared. If we do not use benchmarks (as in all existing prediction systems, including this proposed one) then the values of metrics can be said to be relevant only compared to other values in the graph. For example, we could consider a node with an *outdegree-mean* of nine to be high, if the average node's *outdegree-mean* in the graph is only two. We could also say the same node's *outdegree-mean* as a percentage of the graph as a whole is 450% (i.e. $9 / 2$). This would be judging a person by the similarity of him to his peers using a relative scale. However, even though a node with a mean of 450% may be high for a given particular network, it may be insignificant compared to the mean found in most social networks. Thus the benchmark question presents itself: What is the *outdegree-mean* of most human social networks? Does one exist? Also, does this value, and the universal values for other metrics, differ from group to group, and how so? Do dark networks have different diameter-variances to law enforcement agencies? Do online dating networks have similar indegree-mean values to university email networks? These questions could be answered only by a large-scale study that analysed many different datasets gathered over time from diverse social sources.

Conclusion

In our research we implemented an adaptive agent architecture that functions as a complex adaptive system, based on Marvin Minsky's model of consciousness. The BaBe agent architecture consists of Bayesian agencies and competence agencies. The Bayesian agencies form the A-Brain, observing and learning the structure of CRCs. The competence agencies form the B-Brain – observing and acting upon patterns mined by the Bayesian agencies – thus achieving self-awareness. The competence agencies make CRCs explicit by representing the CRCs using topic maps. We are currently researching the use of the BaBe agent architecture to mine and understand SRCs in order to deal with social complexity.

References

Amit, R. & P. J. Schoemaker (1993). Strategic Assets and Organisational Rent, *Strategic Management Journal*, Vol. 14, pg. 33-46.

April, K. (2002). Guidelines for Developing a K-Strategy. *Journal of Knowledge Management*. Vol 6, No. 5, pg. 445-456.

April, K. & F. Ahmadi-Izadi (2004). *Knowledge Management Praxis*. Kenwyn: Juta Academic.

April, K. & J. Cradock (2000). *e or b e@ten: E-Business Redefining the Corporate Landscape in South Africa*. Durban: Butterworth-Heinemann.

Barney, J. (1991). Firm Resources and Sustained Competitive Advantage, *Journal of Management*, Vol. 17, No. 1, pg. 99-120.

Baas, N. A. & Emmeche, C. (1997). On Emergence and Explanation. *Intellectica*, 25, 67-83.

Barnard, C.I. (1938). *The Functions of the Executive*, Cambridge, MA: Harvard University Press.

Bogaert, I., Maertens, R. & A. Van Cauwenbergh (1994). Strategy as a Aituational Puzzle: The Fit of Components, in G. Hamel & A. Heene (Eds), *Competence-Based Competition*, Chichester: John Wiley.

Burt, R.S. (1992). *Structural Holes: The Social Structure of Competition*, Cambridge, MA: Harvard University Press.

Campbell, C., Maglio, P., Cozzi, A., Dom, B. (2003, November). Expertise Identification using Email Communications. *CIKM*.

Carley, K. (forthcoming). Dynamic Network Analysis. In the Summary of the *NRC workshop on Social Network Modeling and Analysis*. National Research Council.

Coffman, T., Greenblatt, S., Marcus, S. (2004, March) Graph-Based Technologies for Intelligence Analysis. *Communications of the ACM*. 47(3).

Conner, KR. & C.K. Prahalad (2002). A Resource-Based Theory of the Firm: Knowledge versus Opportunism, in C.W. Choo & N. Bontis (Eds.), *The Strategic Management of Intellectual Capital and Organizational Knowledge*, New York: Oxford University Press, Inc., pg. 103-131.

Day, G.S. (1994). The Capabilities of Market-Driven Organisations, *Journal of Marketing*, Vol. 58, pg. 37-52.

Dierickx, I. & K. Cool (1989). Asset Stock Accumulation and Sustainability of Competitive Advantage, *Management Science*, Vol. 35, pg. 1504-11.

Dyer, J.H. & H. Singh (1998). The Relational View: Cooperative Strategy and Sources of Interorganizational Competitive Advantage. *Academy of Management Review*, Vol. 23, No. 4, pg. 660-679.

Fellman, P., & Wright, R. (2004) *Modeling Terrorist Networks - Complex Systems at the Mid-Range*. Retrieved May 31, 2005, from <http://www.psych.lse.ac.uk/complexity/Conference/FellmanWright.pdf>.

Galaskiewicz, J. (1979), *Exchange Networks and Community Politics*, Beverly Hills: Sage.

Granovetter, M.S. (1985). Economic Action and Social Structure: The Problem of Embeddedness. *American Journal of Sociology*, Vol. 91, No. 3, pg. 481-510.

Granovetter, M.S. (1992). Problems of Explanation in Economic Sociology, in N. Nohria & R.G. Eccles (Eds.), *Networks and Organizations: Structure, Form and Action*, Boston, MA: Harvard Business School Press, pg. 25-56.

Grant, R.M. (1996). Prospering in Dynamically-Competitive Environments: Organizational Capability as Knowledge Integration, *Organization Science*, Vol. 7, No. 4, pg. 375-387.

Grant, R.M. (1991). The Resource-Based Theory of Competitive Advantage, *California Management Review*, Vol. 33, No. 3, pg. 114-35.

Hanneman, R. (2001). *Introduction to Social Network Methods*. Retrieved May 31, 2005, from <http://faculty.ucr.edu/~hanneman/SOC157/NETTEXT.PDF>.

Holland, J. H. (1995). *Hidden Order: How Adaptation Builds Complexity*. Massachusetts :Addison-Wesley Publishing Company Inc.

Holme, P. (2003) *Network dynamics of ongoing social relationships*. Retrieved May 31, 2005, from http://arxiv.org/PS_cache/cond-mat/pdf/0308/0308544.pdf.

Holme, P., Edling, C., Liljeros, F. (2004). Structure and Time Evolution of an Internet Dating Community. *Social Networks*. 26. Retrieved May 31, 2005, from <http://www.elsevier.com/locate/socnet>.

Kautz, H., Selman, B., Shah, M. (1997, March) Referral Web: Combining Social Networks and Collaborative Filtering. *Communications of the ACM*, 40(3).

Liben-Nowell, D., & Kleinberg, J. (2003, November). The Link Prediction Problem for Social Networks. *CIKM*.

Lincoln, J.R. (1982). Intra- (and Inter-) Organizational Networks, in S.B. Bacharach (Ed.), *Research in the Sociology of Organizations*, Vol. 1, Greenwich, CT: JAI Press, pg. 1-38.

Loasby, B.J. (1999). Capabilities, in *Knowledge, Institutions and Evolution in Economics*, Chapter 4, London: Routledge, pg. 49-68.

McNee, S., Albert, I., Cosley, D., Gopalkrishnan, P., Lam, S., Rashid, S., Konstan, J., Riedl, J. (2002, November). On the Recommending of Citations for Research Papers. *CSCW*.

Minsky, M. (1988). *The Society of Mind* (First Touchstone ed.). New York: Simon & Schuster.

Nahapiet, J. & S. Ghoshal (1992). Social Capital, Intellectual Capital and the Organizational Advantage, *Academy of Management Review*, Vol. 23, No. 2, pg. 242-266.

Nelson, R.R. & S.G. Winter (1982). *An Evolutionary Theory of Economic Change*, Cambridge, MA: Belknap Press.

Nohria, N. (1992). Is a Network Perspective a Useful Way of Studying Organizations?, in N. Nohria & R.G. Eccles (Eds.), *Networks and Organizations: Structure, Form and Action*, Boston, MA: Harvard Business School Press, pg. 1-22.

N. Nohria & R.G. Eccles (1992). Face-to-Face: Making Network Organizations Work, in N. Nohria & R.G. Eccles (Eds.), *Networks and Organizations: Structure, Form and Action*, Boston, MA: Harvard Business School Press, pg. 288-308.

Pearl, J. (1988). Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference (2nd ed.). San Mateo, USA: Morgan Kaufmann Publishers.

Pearl, J. & Russell, S. (2000). *Bayesian networks, Technical Report R-277, UCLA Cognitive Systems Laboratory*.

Penrose, E.T. (1959). *The Theory of the Growth of the Firm*, 1st Edition, Oxford: Basil Blackwell.

Penrose, E.T. (1995). *The Theory of the Growth of the Firm*, 3rd Edition, Oxford: Oxford University Press.

Popescul, A., & Ungar, L. (2003). *Statistical Relational Learning for Link Prediction*. Retrieved May 31, 2005, from <http://www.cis.upenn.edu/~popescul/Publications/statistical03link.pdf>.

Potgieter, A. (2004). The Engineering of Emergence in Complex Adaptive Systems. *Phd Thesis*, University of Pretoria, Supervised by Bishop, J.

Potgieter, A., April, K. & Bishop, J. (2005). Complex Adaptive Enterprises. in M. Khosrow-Pour (Ed.), *Encyclopedia of Information Science and Technology*, Vol. 1, Hershey, PA: Idea Group Inc., pg. 475-480.

Pujol, J., Sanguesa, R., Delgado, J. (2002, July). Extracting Reputation in Multi-Agent Systems by Means of Social Network Topology. *AAMAS*.

Ross, S., Westerfield, R., Jordan, B., Firer, C. (2001). *Fundamentals of Corporate Finance*, Australia: McGraw-Hill.

Russell, S. J. & Norvig, P. (2003). *Artificial Intelligence A Modern Approach* (2nd ed.). New Jersey, USA: Prentice Hall.

Ryle, G. (1949). *The Concept of Mind*, London: Hutchinson.

Schulze, W.S. (1994). The Two Schools of Thought in Resource-Based Theory: Definitions and Implications for Research, P. Shrivastava, A.S. Huff & J.E. Dutton (Eds), *Advances in Strategic Management*, Vol. 10 A, Greenwich, CT: JAI Press.

Teece, D.J., Pisano, G. & A. Shuen (1990), *Firm Capabilities Resources and the Concept of Strategy* (Working Paper EAP-38, University of California, California, USA).

Taskar, B., Abbeel, P. Wong, M-F., Koller, D. (2004). *Label and Link Prediction in Relational Data*. Retrieved May 31, 2005, from http://kdl.cs.umass.edu/srl2003_upload/files/taskar-paper.pdf.

Van Wijk, R., Van Den Bosch, A.J. & H.W. Volberda (2003). Knowledge and Networks, in M. Easterby-Smith & M.A. Lyles (Eds.), *The Blackwell Handbook of Organizational Learning and Knowledge Management*, Oxford: Blackwell Publishing, pg. 428-453.

Wassermann, S., & Faust, S. (1994). *Social Network Analysis: Methods and Applications*. Cambridge University Press.

Winter, S.G. (1987). Knowledge and Competence as Strategic Assets, in D. Teece (Ed.), *The Competitive Challenge*, New York: HarperCollins, pg. 159-184.