The Net Effect: Modeling and Measuring the Integration of Electronic Communities

Marshall Van Alstyne
Erik Brynjolfsson

Tel: (617) 253-2970
Tel: (617) 253-4319
FAX: (617) 258-7579
marshall@mit.edu
ebrynjo@mit.edu

MIT Sloan School
50 Memorial Drive E53-308
Cambridge, MA 02142

KEYWORDS: Information Economy (AD02), Economic Theory (AM), Economic Impacts (BA01), Computerization of Society (BD0101), Organizational Structure (DA03), Information Flows (DD07), Globalization (AF1301)

ACKNOWLEDGMENTS: This paper has benefited from helpful comments by Mark Ackerman, Brian Butler, Rob Fichman, Rob Kling, David Meyer, Jorge Schement, two anonymous reviewers, participants in the Cambridge Roundtable, the MIT Coordination Science Workshop, a seminar at the University of Michigan, and the International Conference on Information Systems.

An earlier version received the award for Best Paper on the conference theme at the 17th International Conference on Information Systems, Cleveland, OH (Dec. 1996). A significantly shortened precis also appears in the Nov. 29th issue of *Science* (pp. 1479-1480) under the title "Could the Internet Balkanize Science?"

ABSTRACT: Information technology can link geographically separated people and help them locate interesting or compatible resources. Although these attributes have the potential to bridge gaps and unite communities, they also have the potential to fragment interaction and divide groups by leading people to spend more time on special interests and by screening out less preferred contact. This paper introduces precise measures of information integration and develops a model of individual knowledge profiles and community affiliation. These factors suggest different conditions under which improved access, search, and screening can integrate or fragment interaction. As IT capabilities continue to improve, preferences, not geography or technology, become the key determinants of community boundaries.

I. Introduction -- The Emerging Global Village?

With the explosive growth in Internet connections worldwide, networked communication has the potential to shrink geographic distances and facilitate information exchange among people of various backgrounds. Telecommunications policy in the United States -- and other countries -- resolves to extend access to all levels of society, assuming that this will foster greater information exchange while boosting economic growth (NTIA, 1993).

Empowered by information technology such as search engines and automatic filters, IT users are spending more of their waking hours plugged into the Internet, choosing to interact with information sources customized to their individual interests. No longer limited to sources or companions in their geographic neighborhoods, these users presage an interactive world without borders.

What then, are the social and economic consequences of hooking up the next billion users? Does the emergence of a global information infrastructure imply the emergence of the global village -- a virtual community of neighbors freed of geographic constraints?

In this paper, we show that an emerging global village represents only one outcome from a range of possibilities. It is also possible that improved communications access and filtering technologies could further fragment intellectual and social interactions. In particular, we focus on the potential for preferences to reshape social, intellectual and economic neighborhoods as distinct from those based on geography. Just as separation in physical space can divide geographic groups, we find that separation in virtual space can divide special interest groups. In certain cases, the latter can be more fragmented. We introduce several formal indices of integration then show both algebraically and graphically the conditions under which these indices will rise or fall with different levels of access.

The conclusion that increased connectivity and improved filtering could lead to less integration is based on two observations. First, bounded rationality, a limit on the human capacity for calculation (Simon, 1957), can lead to specialization, which decreases the range of overlapping activities. As IT eliminates geographical constraints on interaction, bounded rationality imposes a new constraint. Improved technologies have increased information transmission and bandwidth across all distances except the last 12 inches -- between the computer monitor and the brain. The amount of data one can absorb is bounded, regardless of how fast it scrolls across the screen.[1] Therefore, the choice to attend to new information eventually must also entail the choice to ignore other information. Filters, even sophisticated electronic filters, must be selective in order to provide value.

The second observation is that IT can provide a lubricant that allows for the satisfaction of preferences against the friction of geography. As a result, improving communications and filtering technology helps people increasingly fulfill such preferences. A preference for contact that is more focused than contacts available locally leads to narrower interactions. Thus local heterogeneity can give way to virtual homogeneity as communities coalesce across geographic boundaries. While people often prefer to focus their (finite) information gathering on narrow topics, either because of their intrinsic preferences or because of external rewards to specialization, technology may enable them to indulge this taste to a vastly greater degree than was previously possible. On the other hand, tastes for broader knowledge, or even randomized information, can also be indulged – underscoring how the technology serves mainly to amplify individual preferences.

The same mechanisms that affect the specialization of knowledge also affect the degree to which interactions among people and communities become more or less integrated. The Internet can provide access to millions of other users and a wide range of knowledge sources, but no one can interact with all of them. Even if people wished to do so, creating a global community that depends on individuals consuming vast amounts of disparate and topically unrelated information would simply be infeasible. The practical implication of bounded rationality in this context is that a citizen of cyberspace still has a finite set of "neighbors" with whom he or she can meaningfully interact, but these neighbors can now be chosen based on criteria other than geography. Yet, the number of neighbors with whom one interacts is unlikely to exceed a few dozen in a typical day;

[1] As of May 1996, the AltaVista search engine had indexed more that 33 million articles and web pages. It would take over five years to read just the new listings added each month. This is under the generous assumption that one could access and read a page every 10 seconds, 8 hours a day, 365 days a year and that no pages needed to be revisited as their content changed.

even in a lifetime, few people have significant relationships with more than a few thousand others. As long as human information processing capabilities are bounded, electronic media are unlikely to dramatically change this total. When geography no longer limits interaction, people are able to select their acquaintances by other criteria such as common interests, status, economic class, academic discipline, or ethnic group. In some cases, the result can be a greater balkanization along dimensions that matter far more than geography, while in other cases more diverse communities may emerge. However, our analysis suggests that automatic search tools and filters that route communications among people based on their views, reputations, past statements or personal characteristics are not necessarily benign in their effects.

Of course, preferences themselves need not remain unaffected by such tools. Because the Internet makes it easier to find like-minded individuals, it can facilitate and strengthen fringe communities that have a common ideology but are dispersed geographically. Thus, particle physicists, oenophiles, Star Trek fans, and members of militia groups have used the Internet to find each other, swap information and stoke each others' passions. In many cases, their heated dialogues might never have reached critical mass as long as geographic separation diluted them to a few parts per million. Once like-minded individuals locate each other, their subsequent interactions can further polarize their views or even ignite calls-to-action. The Internet can also facilitate the de facto secession of individuals or groups from their geographic neighborhoods. Because time is limited, spending more time interacting with online communities necessarily means spending less time interacting with geographic communities or even family members. For instance, while Jose Soriano founded a Peruvian network "to minimize the gap between the information haves and information have-nots," the network also facilitates local geographic secession. Psychologist Manuel Molla Madeueno, a network user, reports that "I use the Internet to read psychology magazines and articles and notes that are posted on the psychology bulletin board. The only problem is that I've become obsessed with what I can do on the Internet and I'm spending all my free time there" (Sims, 1996). As Mr. Madeueno becomes more of a member of the community of academic psychologists, he inevitably becomes less of a member of some other community such as his Peruvian village, at least in terms of time spent interacting. The Internet has apparently led him to spend less time interacting with his geographic neighbors, isolating him on some dimensions even as it integrates him on others.

We do not argue that increased specialization or fragmentation must result from increased connectivity. On the contrary, we believe that the Internet has enormous potential to elevate the nature of human interaction. Indeed, we find that if preferences favor diversity, increased connectivity reduces specialization. However, our analysis also indicates that, other factors being equal, all that is required to reduce integration is that preferred interactions are more focused than existing interactions. A desire for increased focus and improved filtering of noisy communications is a natural response to data and computational overload. Although the conventional wisdom has stressed the integrating effects of the technology, we examine critically the claim that a global village is the inexorable result of increased connectivity.

Existing literature provides useful indices of "centrality" and "vulnerability" in network structures (Alstyne, 1997; Freeman, 1979; Malone & Smith, 1988) which help measure how well-connected or pivotal individuals are within their networks. Social network literature also identifies blocks of related ties within interconnected communities (White, Boorman & Breiger, 1976) and describes the effects of weak ties (Granovetter, 1973) or the absence of ties i.e. "structural holes" (Burt, 1993) on information flows. A formal model of dyadic or pairwise communication also shows integration occurring as a result of face-to-face interaction and print communication (Kaufer & Carley, 1993). Complementing this literature, our research provides a model of shared information and community cohesion when multiple simultaneous interactions are possible. It also provides specific new measures of integration. We use these indices to examine theoretical implications of changing interconnectivity, searching, and screening.

II. Modeling People and Resources (Measures of Integration)

To examine community interactions, we construct a model of individual contact and information resource distribution. By "community," we mean groups of individuals that participate in joint contact and information sharing and we acknowledge that there are broader conceptualizations that include, for example, companionship and emotional support (Wellman & Gulia, 1997). Then, since "integration" lends itself to several interpretations, we introduce multiple measures including: the overlap or level of contact between groups, the directness or distance of communication, and the level of concentration in information resources.[2] Let the agents be enumerated as $i, j \in \{1, 2, 3, \ldots \mathcal{N}\}$ where $\mathcal{N}$ is the size of the total population. Then we can say that access $\mathcal{A}$ improves as it increases from 1 to $\mathcal{N}$ and that $\mathcal{A}/\mathcal{N}$ represents the fraction of the population any given agent i can potentially reach. Also, each agent has C channels, the maximum number of people from the population he or she can contact simultaneously assuming bounded rationality. For example, the telephone network might grant one access to billions of people ($\mathcal{A}$) but time constraints might permit interaction with no more than several dozen in a day (C).

Adopting the convention of an information resource as a knowledge base represented by k_{it}, we can associate knowledge with individual agents i in terms of both a type $t \in \{1, 2, 3, \ldots \mathcal{T}\}$ and relative amount k.[3] Importantly, this also allows us to distinguish access by type and to characterize knowledge profiles by agent. Let the knowledge profile $\mathcal{P}_i$ of agent i be a vector of how much he knows about each topic $\mathcal{P}_i = [k_{i1}, k_{i2}, \ldots k_{i\mathcal{T}}]$. Each agent can thus be mapped to a unique point in "knowledge space" which is analogous to his or her geographic location. If an agent starts with only a single type of information and has knowledge profile $\mathcal{P}_i = [0, 0, \ldots k_{it}, \ldots 0]$ then allowing access to an agent j who has knowledge of a different topic s can potentially provide agent i with a profile of $\mathcal{P}_i = [0, 0, k_{js}, k_{it}, \ldots 0]$. Then, if k_t is the total knowledge of a given type i.e. $k_t = \sum_{i=1}^{\mathcal{N}} k_{it}$ we can describe the total knowledge existing in a population as $\mathcal{K} = [k_1,$

[2] A table of precise interpretations for our constructs appears in a glossary of symbols at the end of this article.
[3] In this paper, we assume a cardinal measure of knowledge exists. For most analysis, a simple boolean value, 0 or 1, will suffice with a commensurately larger number of knowledge types i.e. either a fact is known or it isn't.

k_2, ... $k_{\mathcal{T}}$]. For simplicity, we do not require agents with the same type of knowledge to know exactly the same information. Thus agents with overlapping information can connect with a net gain in resources. Under these assumptions, increasing access has the attractive property of increasing an agent's knowledge profile towards full information where $\|\mathcal{P}_i\|/\|\mathcal{K}\| = 1$. The magnitude of the knowledge profile indicates how close an individual agent comes to accessing the full information available to a society of individuals.

Shared Knowledge Index: Using this terminology, we now have the ability to calculate the degree of "similarity" between knowledge profiles $\mathcal{P}_i$ and $\mathcal{P}_j$ represented as the cosine of Θ_{ij} the angle between them.

Definition: The "similarity," $\mathcal{S}_{ij}$, between two individuals in "knowledge space" is given by: $\mathcal{S}_{ij} = \mathrm{Cos}(\Theta_{ij}) = \mathcal{P}_i \cdot \mathcal{P}_j / \|\mathcal{P}_i\| \, \|\mathcal{P}_j\|$.

$\mathrm{Cos}(\Theta_{ij})$ approaches 1 as profiles become more similar and approaches 0 as they grow farther apart.[4] Occasionally, it will also be convenient to index a group of agents by the overall average of knowledge profiles $\frac{1}{\mathcal{N}^2}\sum_i\sum_j \mathrm{Cos}(\Theta_{ij})$.

Shared Knowledge Distance: An alternative to $\mathcal{S}_{ij}$ is to compute the "distance" D_{ij} between knowledge profiles by applying a distance metric to their difference. This identifies how many of the knowledge bases agents share.

Definition: The "distance" between two individuals *i* and *j* in "knowledge space" is:

$$\|\mathcal{P}_i - \mathcal{P}_j\| = \sqrt{(k_{i1} - k_{j1})^2 + (k_{i2} - k_{j2})^2 + \cdots (k_{i\mathcal{T}} - k_{j\mathcal{T}})^2}$$

If both agents *i* and *j* have access to exactly the same knowledge bases[5] then this expression reduces to zero but this index can otherwise take on values in $\mathbb{R}^+$.

---

[4] In this paper, we assume that "negative" knowledge does not exist, thus $\mathrm{Cos}(\Theta_{ij})$ is non-negative.

[5] If we wish to allow for the possibility of knowledge overlap, then type differences become set differences e.g. $|k_{it} \cup k_{jt} - (k_{it} \cap k_{jt})|$. Similar changes in other indices provide consistent results.

Integrated Affiliations: Based on these profiles, we can also define an index of how much agents' group memberships or affiliations overlap. An agent, who starts out with resources of type t_1, can increase affiliations by gaining access to other types t_2. For an index of affiliation, we want a measure that increases when communities overlap and that decreases with the number of separate communities. Let the members affiliated with a community of type t be given by $\mathcal{M}(t)$ so that we can derive a metric of group overlap.

Definition: The "index of integrated affiliation," μ_A for a population is given by:[6]

$$\mu_{\mathcal{A}} \equiv \frac{1}{\mathcal{T}} \frac{1}{\mathcal{T}-1} \sum_{t \in \{1,2\ldots\mathcal{T}\}} \sum_{s \neq t} \frac{\|\mathcal{M}(t) \cap \mathcal{M}(s)\|^2}{\|\mathcal{M}(t)\| \|\mathcal{M}(s)\|}$$

This index ranges from a low of 0 if every community is closed and shares no members in common with any other community to a high of 1 if every individual is a member of every community. The more diverse an agent's particular affiliations, the more he or she raises the index of integrated affiliation.

Integrated communication: This index measures the integration of channel paths or who talks to whom. In a fragmented community, agents communicate in clusters or possibly not at all. In a fully integrated community, each agent communicates with everyone. For an index of integrated communication, we require a measure that decreases in the number of isolated agents and that increases each time agents establish a connection. If agents are connected in a graph, let the communication distance between two agents i and j be the total number of links $\mathcal{L}_{ij}$ on the shortest path between them. Note that $\mathcal{L}_{ij}$ need not equal $\mathcal{L}_{ji}$ if communication is directional or agents use different intermediaries. Also, since agents do not need to connect to themselves, the least upper bound on a chain of connections among $\mathcal{N}$ agents is $\mathcal{N}$-1. If no chain of connections exists between i and j, define the distance to be $\mathcal{N}$.

Definition: With these terms, we may define μ_C, the measure of integrated communication, as:

[6] This index is a generalization of a measure of two-way overlap appearing in (Donath, 1995).

$$\mu_c \equiv 1 - \frac{1}{\mathcal{N}^2} \sum_{i \in \{1,2 \dots \mathcal{N}\}} \sum_{j \neq i} \frac{\mathcal{L}_{ij}}{\mathcal{N} - 1}$$

Thus $\mu_C \in [0, 1)$, approaching 1 (i.e. $1-\frac{1}{\mathcal{N}}$) as the population becomes large for highly integrated topologies. It reaches its lowest value when every agent is a single disconnected island and its highest value when every agent is directly connected.

Integrated Resources: The degree to which knowledge bases are concentrated can vary independently of whether specific agents are directly connected i.e. whether communication itself is integrated. A refusal to share, for example, would balkanize information despite the existence of a channel whereas access via an alternate source would integrate the same resource. For an index of integrated information, we require the measure to decrease as more resources become inaccessible to any single agent and to decrease also as the number of agents find the same resource inaccessible.

Definition: Using our constructs for knowledge, we define the index of integrated information μ_I to be

$$\mu_{\mathcal{I}} \equiv \frac{1}{\mathcal{N}} \frac{1}{\mathcal{T}} \sum_{i \in \{1,2 \dots \mathcal{N}\}} \sum_{t \in \{1,2 \dots \mathcal{T}\}} \left(\frac{k_{it}}{k_t} \right)^2$$

This index ranges from 0 when a single agent has exclusive access to all of a society's knowledge resources $\mathcal{K}$ to a maximum of 1 when the entire population has access to $\mathcal{K}$. Although we base this index on information shares, μ_I could equally well be used to measure other resource concentrations.[7]

The similarity measure $\mathcal{S}_{ij} = \mathrm{Cos}(\Theta_{ij})$ provides an index of comparative individual access. The integration measure μ_A indexes the diversity of group interactions among members of a

[7]This measure adapts the Herfindahl index of market share concentration to multiple dimensions. This index is similar in spirit to a Gini index, which measures concentration as the sum of deviations from the 45 degree line on the plot of sorted resources versus sorted individuals. If the botton 25% of the population still earned 25% of all income, for example, then income plots at 45˚ and is equitably distributed.

society. The two additional measures complement these two; μ_C refers to communications and μ_I to information resources.[8] Although they can move independently, in most cases, the results tend to be qualitatively similar so we focus on $\mathcal{S}_{ij}$ and $\mu_{\mathcal{A}}$. This collection of indices provides a way to compare both individuals and groups within a society based on the same constructs of information access and affiliation. An example of $\mathcal{S}_{ij}$, $\mu_{\mathcal{A}}$, μ_C, and $\mu_{\mathcal{I}}$ is provided in the next section.

III. Geography Unbound

As communication costs fall generally, the cost of connecting individual agents also falls. If the costs are too high, no two agents communicate; if the costs are negligible, all agents can communicate. With IT costs falling dramatically, inter-connectivity is likely to increase [Malone, 1987 #53; Wyner, 1996 #452]. One possible progression is a move from completely isolated agents to completely interconnected agents as in Figure 1. We use these to illustrate the indices of integration.

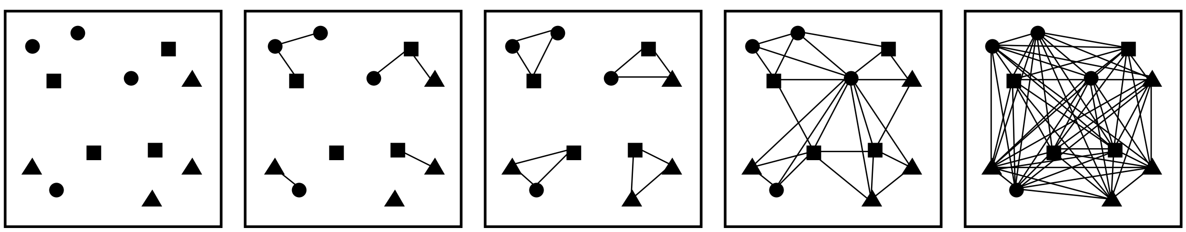

Figure 1.A **Figure 1.B** **Figure 1.C** **Figure 1.D** **Figure 1.E**

Connectivity levels increase as communication costs fall from left to right.

This example conforms to popular ideas on the emergence of networked infrastructure. When communication costs are prohibitive, these twelve agents operate in isolation with incomplete knowledge of global information as in Figure 1.A. As communication costs fall, clusters of communication emerge allowing agents to share information and gain a less fragmented understanding. This is shown in intermediate frames. Once costs become negligible, a fully connected community emerges permitting everyone access to full knowledge of events as in Figure 1.E. From left to right, knowledge profiles grow from their greatest fragmentation to their least

[8] Note that indices of "balkanization," β, could just as easily be represented as 1-μ.

fragmentation while community "integration" increases. Different agents, represented by different shapes may have different information requirements or communication interests. These potential preferences will motivate subsequent observations on how much communication actually occurs. For instance, even when agents are *able* to communicate with all other agents, they may not be willing to do so. For now, we assume that agents are both willing and able so the potential of increased access is, in fact, realized. The basic intuition is shown formally in Proposition 1.

Proposition 1 -- Without bounded rationality constraints, global access maximizes integration when agents do not block communication. That is $\mu_A = 1$ and agents' knowledge profiles are the same, $\mathcal{S}_{ij} = 1$.

Proof: Details of each proof are provided in an Appendix.

The table below shows the data for $\mathcal{S}_{ij}$=Cos(Θ_{ij}) and for μ_A using the graphs from Figure 1.

Index	Figure 1.A	Figure 1.B	Figure 1.C	Figure 1.D	Figure 1.E
Average($\mathcal{S}_{ij}$)	.27	.55	.77	.84	1
$\mu_{\mathcal{A}}$	0	.25	.65	.83	1
$\mu_{\mathcal{C}}$	0	.11	.17	.87	.92
$\mu_{\mathcal{I}}$	.02	.05	.08	.17	1

Table 1 – Measures of integration change with the communication changes of Figure 1.

In this example, there are 4 agents of each type so their knowledge profiles overlap somewhat in Figure 1.A. If there were 12 separate types, the similarity measure would be 0. By Figure 1.E, all agents have access to society's information so knowledge profiles are identical. Communities of types in Figure 1.A, however, share no members in common so μ_A indicates complete segregation. Once the types are completely interconnected, this index rises to 1. The other metrics for communications distance $\mu_{\mathcal{C}}$ and information concentration $\mu_{\mathcal{I}}$ are provided for illustration.

The rise in integration associated with improved access in this simple model is consistent with the common view that telecommunications, and the Internet in particular, can help foster the emergence of a global village.

IV. Rationality Bound

The trouble with simply eliminating geographic constraints lies in assuming the absence of any other constraints such as bounded rationality or vetoed interaction. Physical connectivity does not necessarily imply logical connectivity when either party at one end of a connection is either too preoccupied or otherwise unwilling to interact. In practice, limitations on interaction exist due to (1) cognitive capacity constraints e.g. e-mail overload (2) missing or unshared vocabulary, e.g. medical terminology (3) insufficient bandwidth, e.g. even video-conferencing may provide insufficient context for first meetings and (4) lack of trust, e.g. certain business relationships may require bonding time to establish reputations.[9]

Since time and human information processing capacity are finite, agents prioritize their connections based on a combination of access to and preferences for certain types. The figure below depicts expanding geographic access. Nearer agents are more likely contacts as are agents of a preferred type.

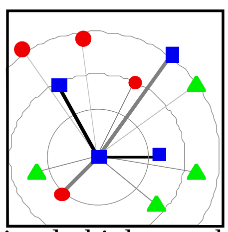

Figure 2 -- Darker thicker lines imply higher probabilities of a connection viewed from the perspective of a type at the center of expanding access circles.

As access expands, more agents become possible contacts. The agent shown near the center prefers to connect to like types if they happen to be within reach. Characterizing this yields Proposition 2.

[9]Unconstrained communication can actually be burdensome. During one police investigation, an Internet posting of a request for information resulted in too many false leads during a time-sensitive abduction (Leslie, 1995). Netiquette also addresses wasteful communication. Newsgroup readers actively discourage posting irrelevant material -- off-topic news, solicitations, personal attacks -- partly because of the time and nuisance costs it imposes on the community.

Proposition 2 -- Virtual communities decrease integration relative to geographic communities given bounded rationality, $\mathcal{C}<\mathcal{T}$, and a preference for similar[10] types.

The figures below illustrate geographic connections converting to type connections after agents gain global access.

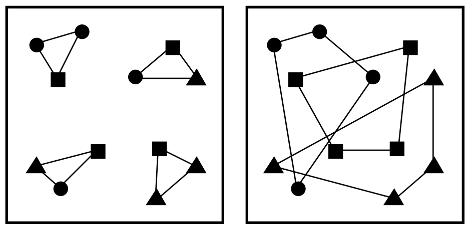

Figure 3.A **Figure 3.B**

As geographic access improves, agents seek preferred types.

In Figure 3.A, access is strictly local and bounded by geography so agents form small communities with fairly uniform knowledge access across types. Figure 3.B shows global access; like types have located one another and formed tightly knit communities. Interconnections between communities have fallen and resources are more concentrated. The integration indices in this example show that for twelve agents and three types, the average $\mathrm{Cos}(\Theta_{ij})$ declines from .77 to .27 indicating that profiles diverge, and μ_A falls from .65 to 0, indicating that overlap has fallen among communities.

Except in highly segregated societies, geography imposes an unavoidable heterogeneity. Adam Smith's butcher, brewer, and baker rubbed shoulders in a local town of physical neighbors. A virtual community of like-minded citizens, however, might be virtually homogeneous. The baker might learn from and contribute to the collected information on baking, indulging his passion for learning more and more about breads even as he learns less and less about meat and drink. As virtual citizens leave their physical neighborhoods behind, they inadvertently withdraw their contributions to their physical locales. Increasingly unaware of each other's communities,

[10]In these examples, we follow Kaufer and Carley (1993) and show preferences as an affinity for similar types. This is primarily helpful as illustration and is inessential to our main findings. Later examples will show how preferences for dissimilar types can give analogous results.

specialized vocabularies, and contributions, the butcher, the brewer, and the baker might each provide better meat, drink, or bread individually but lose the ability to provide a balanced meal.

Reduced integration is by no means inevitable, however. Incentives can coax groups into working together. If there are benefits to cooperation, then IT is likely to facilitate interaction among them. A caterer, for example, might contract with providers from each community. Geographic constraints removed, incentives rather than random chance will determine which interactions occur, which raises the issues of how incentives operate and of which pathways emerge.

Evidence suggests that better communications can strengthen interactions within special interest groups. In describing what might characterize the emerging "Global Village," Marshall McLuhan and Powers (1989) nevertheless recognize the power of satellite technology to aid "super-regionalisms" and "separatisms" like the Parti Quebeçois in Canada. As an historical example, the telephone strengthened affiliation among teenage peer groups (Sproull & Kiesler, 1991). In the field of economics, the number of out-of-state and out-of-country coauthorships in four top journals grew from 4.6% in the 1960s to 27.6% in the 1990s (Gaspar & Glaeser, 1996). Communications scholars have suggested that virtual communities are often more heterogeneous in terms of social attributes (age, income, gender, ethnicity) but more homogenous in terms of attitudes (Hiltz & Wellman, 1997). Similar evidence appears to hold for academia in general:

> "Historically, the strength of an academic department rested with its resident faculty. Now it depends on the extent to which each faculty member is interconnected with other professionals -- worldwide -- pursuing similar interests... We now have electronic research teams and electronic water coolers. This drastically changes -- weakens, in my opinion -- indigenous workplace relationships and affects workplace cohesiveness."[11]

Contact expansion is also a factor that might influence integration. The Internet has the attractive property that a person might communicate with a very large group of associates, and this might integrate the population. For example newsgroups and chat rooms have few, if any,

---

[11]Interview with Edward Mabry, communications researcher at University of Wisconsin, Milwaukee (Leslie, 1995).

physical limitations on participation. Broader participation will increase the integration of resources and groups insofar as more people choose to access the same information and insofar as information flows across the boundaries of distinct groups. These represent important gains from increased connectivity. There are, however, two important qualifications.

The first qualification is that if broader participation is restricted to members of the same community then information resources are less fragmented (i.e. μ_I rises) but intergroup fragmentation remains unchanged (i.e. μ_A stays constant). People might also find that the sizes of their groups increase but that the number of group memberships they hold remains relatively constant. By analogy to journal publication, the variety of choices and the number of subscribers might increase but the number of simultaneous subscriptions held by individuals might not increase much. Due to bounded rationality, the median subscriber base might even fall as mass publication gives way to niche publication. This appears to be a strategy of many increasingly focused, or even personalized, net "e-zines" and news feeds.

Second, participating in a group is not the same as interacting with all of a group's members. The number of subscribers to most newsgroups substantially exceeds the number of people who actually post messages. Joining a new community shortens communications distance (i.e. μ_C rises) only if this establishes new and shorter paths and it homogenizes information profiles (i.e. μ_I rises) only if members actually communicate. Universal participation is unlikely to imply universal broadcast. Noise and confusion would likely result. Expressed differently, as a group gets larger, the fraction of members who post communications probably declines after some critical point. In network organizations, for example, sociologists have recognized that as the number of in-group ties increase, the number of out-group ties tends to decrease due to affinity relations and economizing on time and effort (Baker, 1993) -- an empirical finding which supports the possibility of fragmentation.

We consider capabilities enhancement a second order effect of the Internet relative to the first order effect of improving communications reach and search capability. Although bounded rationality constraints motivate Proposition 2, expanding capabilities to relax this constraint can leave intact the basic result of decreased integration, as long as reach and search capabilities improve more rapidly. Graphically, this can be shown in the following manner.

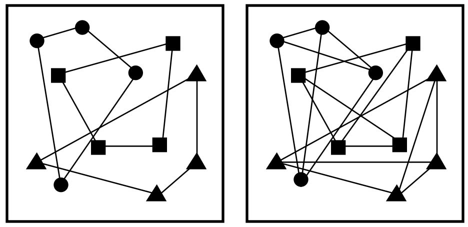

Figure 4.A **Figure 4.B**

Increasing channel capacity does not necessarily increase the measure of integrated affiliations.

In this example, agents use their additional connections to reach additional members of the same community.[12] This leads to the following observation:

Proposition 3 -- Under global access, relaxing the bounded rationality constraint C does not reduce balkanized affiliation if agents choose to veto certain communications or if agents choose not to seek information outside their original topic areas. Let μ_A' and Θ_{ij}' represent new indices after expanding agents' capabilities. If agents connect to members of the same community, then new indices are unchanged and $\mu_A' = \mu_A$ and $\forall\ i, j$, $\Theta_{ij}' = \Theta_{ij}$.

New channels could increase integration if agents used their added capabilities to reach outside their original communities. Again, preferences are crucial. In the example above, however, channels have increased by 50% but all agents use their additional resources to communicate with previously unreached members of their existing communities. Resource fragmentation falls (μ_I changes from .19 to .33) but the index of balkanized affiliation, μ_A, remains unchanged. Each agent effectively deepens his or her knowledge of a given topic area, rather than broadening it to other topic areas.

12This is not to suggest that nothing has changed. Communication paths are (slightly) shorter and information is less (slightly) concentrated implying that β_C and β_I have fallen (xxx note measures!). This draws attention to the importance of multiple measures. The Appendix presents an example in which indices even move in opposite directions.

V. The Preferences That Bind

Another feature of Proposition 2 is that agents exhibit fairly strong preferences: they prefer to associate exclusively with agents of like types. Our fourth result relaxes this condition (some agents might prefer *intentional* randomness) and shows that even weak preferences can lead to similar results. In fact, unless agents are indifferent to their connections or seek greater diversity than is locally available, a population with global access will generally increase on measures of balkanization. We formalize this below.

Proposition 4 -- Narrower preferences reduce integration. If an agent prefers less diverse associations than an average sample from the local population, then integration decreases. Stronger preferences lead to greater balkanization.

In Figure 5, we hold access constant at a global level and vary preferences. Initially, agents enjoy connecting with everyone -- there is an equal (2/3) probability that a connection reaches any type. In the second frame, agents have committed one of their channels to a like type leaving a smaller (1/3) chance of connecting on the remaining channel. In the third frame, agents have allocated both channels to like types reducing the chances of inter-community interaction to zero.

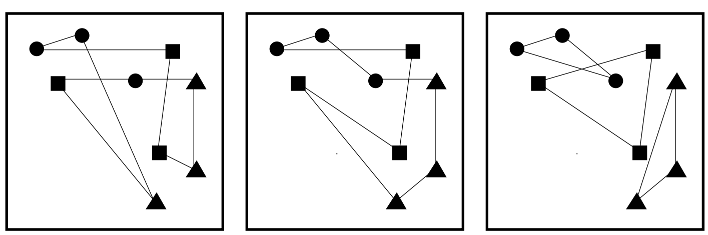

Figure 5.A **Figure 5.B** **Figure 5.C**

Under global access, integration decreases as agents narrow their preferences.

Index	Figure 5.A	Figure 5.B	Figure 5.C
Average(S_{ij})	0.69	0.44	0.25
μ_A	0.51	0.16	0

Modeling this outcome depends on the ease of long distance dialog and the loss of serendipitous interaction, two features which may be increasingly in evidence today. In ruling against certain provisions of the Communications Decency Act, for example, the presiding judges stated in their findings of fact:

> ... unlike traditional media, ... communications over the Internet do not "invade" an individual's home or appear on one's computer screen unbidden. Users seldom encounter content "by accident."[13]

If individuals can choose their content, contacts, and connections, then emphasizing preferred communities can balkanize interactions. Although Figure 5 shows only one type of commitment, this effect does not depend on a preference for a single type. If, for example, a person chooses to interact with a dozen communities when serendipitous geographic interaction would have led to several dozen, then the breadth of exposure to novel types of information is likely to fall.[14] The key point is that if the distribution of tastes differs from the distribution of types in the local population, then technologically-lowered search costs allows people to shift their interactions towards indulging their tastes, possibly lowering levels of integration. With bounded rationality, any preference for membership in virtual communities that is more concentrated than representative geographic samples will reduce affiliations with less favored communities -- the stronger the preferences, the greater the fragmentation. Proposition 4 has powerful implications because it suggests that geography only needs to be more heterogeneous than tastes in order for the lifting of geographic constraints to result in more specialized interactions.

Specialization and fragmentation as outcomes, however, are by no means certain. If tastes favor diversity then the action of preferences can reduce specialization and increase integration. In other words, if geography is less heterogeneous than tastes, greater connectivity implies moving from right to left in Figure 5. This gives rise to the following observation:

Corollary 4 -- If agents exhibit a taste for diversity or randomness, which exceeds that available locally, then greater connectivity leads to more integrated interaction.

---

[13]ACLU v. Reno #96-963. Ruling of 6/11/96; §88.

[14]Nor does the result depend on a preference for *similar* types. If preferences are such that A will only speak to B who will only speak to C who will only speak to A, then communications deadlock. Likewise, mutual preferences for contact between pairs (A, B) and (C, D) can divide the community into two blocks. In both cases, a preference for dissimilar types can balkanize interactions.

This corollary simply restates the previous conclusion with tastes running in the opposite direction. The Internet can, in fact, lend itself to experimentation as when people enter into multi user-domains (MUDs) and alter their virtual personalities, genders, and behaviors (Turkle, 1995). Differences can also attract (Kaufer & Carley, 1993). The model can illustrate movement both towards and away from integration. Which effect dominates depends on the aggregate distribution of preferences. Narrower preferences lead to specialization while broader preferences lead to integration, as connectivity increases.

VI. Endogenous Preferences: Hyper-specialization?

Additionally, preferences need not remain static; the desire for affiliation with a particular group can increase or decrease over time. With geographic constraints, the degree of heterogeneity, noise, and serendipity implies that actual interactions have the more diverse and jagged profile shown in Figure 6.A. An agent with access to IT can more easily match a desired profile by searching for people with similar interests, by using searching and filtering to suggest relevant information and by using IT to screen out less attractive partners or information.

IT initially aids search and filtration as in Figures 6.A to 6.B, but agents may subsequently acquire new tastes or sharpen their preferences. The progression from Figure 6.B to 6.C illustrates the condition of positive feedback: an affinity for a particular topic leads an agent to seek either more information from or more solidarity with the community focused on that topic. Communications research suggests that "what you know depends on whom you know and who you know depends on whom you meet" (Sproull & Kiesler, 1991 p.11). This can make knowledge profiles path dependent and it also opens the door to information feedback. Historically, positive feedback is damped by geography and unfiltered interaction. Communication technology, search engines, and message filters, however, support positive feedback.[15] With

[15] Less focused interaction, or negative feedback, might also occur if communities experience satiation. If their interests become more diffuse over time, the progression could move from right to left in Figure 6.

"perfect" filtering and positive feedback, the only stopping point is a topic singularity as appearing in Figure 6.C.

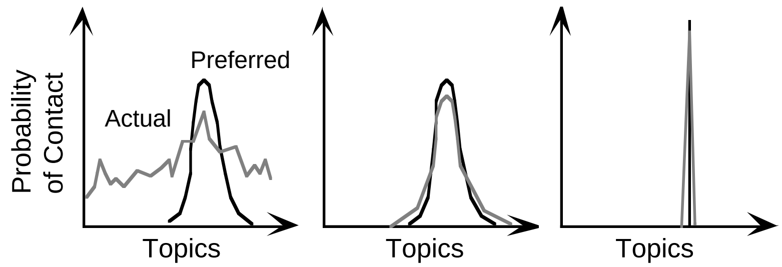

Figure 6.A **Figure 6.B** **Figure 6.C**

Reducing noise and positive feedback can concentrate contact.

The initial effect of IT might be to reduce noise, moving contacts from 6.A to 6.B, while the subsequent effect could be to encourage specialization, moving contacts from 6.B to 6.C. The second shift might represent a political dabbler becoming focused on a special interest group or an oenophile graduating from a general interest in wines to a preference for fine burgundies only. Empirical findings appear to support these observations. In illustrating a theory of communication ecology as mutually defining agents, context, and transmissions, Kaufer and Carley (1993) cite several studies in which contact and shared information exhibited a reciprocal relationship. A study of writing students, for example, found a positive correlation between their patterns of interaction and emerging pockets of consensus. Another study found that employees' similarity, shared information, and proximity predicted their social interactions. Feedback between interaction and shared information also appears to explain differences in observed cohesiveness of certain religious groups (Kaufer & Carley, 1993).

Recent technology provides a far higher level of control over interaction and contact filtering. Control exists at the message level where previously it extended only to the interpersonal level. Moreover, filtering and screening may take place on behalf of individuals with or without their foreknowledge and consent. Several news organizations deliver customized news via web browsers and a patent has recently been issued for software which customizes personalized advertisements over cable channels.[16] Advertising and news stories can be targeted to the level of word choice both to spark interest and to penetrate filters designed to screen unwanted contact. IT might therefore lead to more focused interaction through the action of choices people make for themselves and the action of choices others make for them. Moreover, if preferences for new information are a function of past information or contacts, then an agent's preferences and connections can become highly path dependent. Small changes early in the evolution of a community can radically affect their character.

Many communities offer examples of increasingly narrow focus and specialization. Numerous (most?) academic disciplines, in fact, have progressed to the point where the specialized vocabulary that facilitates interactions within the community hinders interaction across communities. Indeed, Kuhn has observed that a widening gulf "separates the professional scientist from his colleagues in other fields" (Kuhn, 1970, p. 21). Specialists in branches of mathematics other than algebraic geometry, for example, have difficulty following a recent proof of Fermat's last theorem. Differentiation and inbreeding among communities has progressed to the point where expertise can mean “knowing more and more about less and less.” As IT improves filtering, tailoring, segmenting, and searching, the more global network becomes the less local village.

VII. From Specialization to Stratification

The principles which govern integration across types also hold within types if sufficient differentiation exists to distinguish one classification from another of the same type. Quality can represent one such axis of differentiation. Given quality differences -- analogous to different types -- the same potential for fragmentation can occur. Associations can form among high, middle, and low quality tiers inducing stratification. As a twist on specialization, this leads to our final proposition.

Proposition 5 -- Quality differentiation in virtual communities can lead to stratification.

---

[16] US Patent 5,515,098.

In the figures below, icons represent a single type in which superior resources interact with one another.

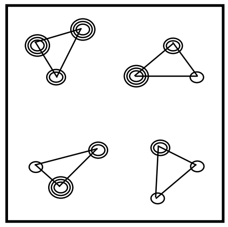

Figure 7.A

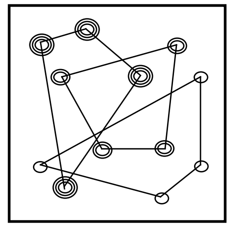

Figure 7.B

These figures, which are analogs of Figures 3.A and 3.B
show stratification within one type.

The same community affiliation mechanisms apply. This interpretation emphasizes one novel dynamic, however, implicit in earlier discussions. Agents at a source might wish to affiliate with agents at a destination but if agents at the destination have already committed their channels, the destination community is closed. Veto power at a destination can balkanize communities despite preferences for diversity at a source. The publisher of an electronic newsletter, for example, argues

> I hate to sound undemocratic, but if you're going to have valuable discussion, you have to limit it to people with valuable knowledge. The beginners can have their beginner's groups. (Chao, 1995).

New information technologies can therefore exclude as well as enfranchise as with encryption, intranets, private mailing lists, and firewalls. With respect to quality, agents might wish to connect to others agents' higher grade resources but find no connections available. Agents in the top tier of an information pyramid might therefore benefit disproportionately from global access. Any initial advantage from owning a quality resource grows because it gains an agent possible entry into an otherwise closed community.[17] Importantly, however, closure occurs not from any bias against one group but from a preference for another. There is an implicit competition between choices for interaction only because a limitation on choices means that a decision to include one option results in the exclusion of another.

---

[17]For a different model of stratification, see (Alstyne & Brynjolfsson, 1995)

The combination of stratification and specialization are depicted in Figure 8. Members of communities A and B might be geographically close (8.A) but have knowledge of diverse topics (8.B). Global networks enable new communities A' and B' which are geographically diverse but specialize with respect to topic area (8.B). Finally 8.C shows how adding other axes of differentiation creates an opportunity for further fragmentation, specifically, what we call stratification. Such axes can include, for example, subspecialization within an existing discipline.

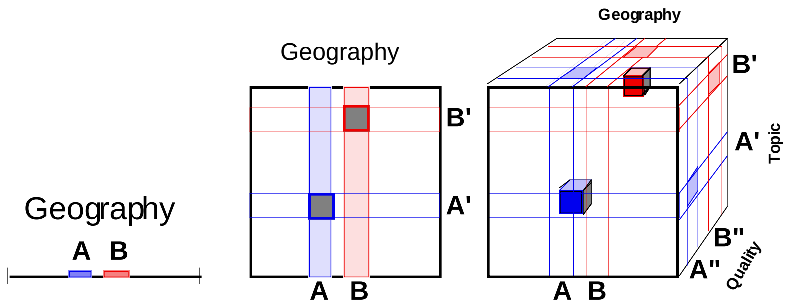

Figure 8.A **Figure 8.B** **Figure 8.C**

Geography's influence on near neighbors diminishes as other dimensions such as topic and quality come to the fore. Fragmentation can increase with each new dimension, progressing from geographic separation (8.A), to topic specialization (8.B), to quality stratification (8.C).

Focusing interaction can homogenize *intra*-action but at the cost of separating groups. Figure 8.C, in particular, shows how one model of community interactions could place near geographic neighbors in different topic camps. If IT shrinks distance, spending time "abroad" can imply neighbors might become strangers and measures of knowledge overlap might fall with rising fragmentation of communities.

VIII. Conclusions & Implications: Should We Care About Fragmentation?

In this paper, we have defined measures of integration, we have developed a model of possible affiliation based on individual preferences, and we have used these tools to explore possible theoretical implications of changes wrought by IT. In particular, these changes affect our capacity to select, search, screen, and connect. As these abilities influence knowledge profiles and community membership, they also influence the diversity and integration of the communities we voluntarily form. Our findings are suggestive of possible future scenarios and of paths dependent on the interactions we choose for ourselves. In this context, conditions set forth in our various propositions may help to guide the choices we might wish to make in light of such factors as diversity, equality, and efficiency. Furthermore, the metrics we define can be used both to advance theoretical analyses and as the basis for empirical assessments of the effects of electronic connectivity.

Under certain conditions, specialization and fragmentation can be economically efficient and stable in the sense that no individual can be made better off by changing their personal affiliations from those under focused interaction. It might even be possible for this to be welfare maximizing in the sense that specialized production may produce the greatest output. Interconnected "collaboratories" can allow scientists to share data and access remote instruments (Wulf, 1993). Oceanographers who confer through communications technology are more productive -- they author more papers, earn greater peer recognition, and know more colleagues (Hesse, Sproull, Kiesler & Walsh, 1993). And, economists sometimes assert that matching peers is efficient (Kremer, 1993; Roth, 1984), as the benefits of specialization date back to the work of Adam Smith (1776).

Voluntary fragmentation, however, might also prove destructive to the overall welfare of society due to over-specialization. Watson and Crick, for example, combined skills from zoology and x-ray diffraction to determine the structure of DNA (Moffat, 1993). Thomas Kuhn developed his ideas on scientific paradigm shifts while working at the nexus of history and physics; yet it would be extremely difficult to look for common principles across paradigms by examining a single paradigm. Once Black and Scholes recognized their formula for options pricing as a physics equation for heat transfer (Black & Scholes, 1973, p. 644) they could look for established parallels. Similarly, the Alvarez theory that an asteroid caused the extinction of the dinosaurs emerged from a fortuitous combination of father and son skills in astrophysics and geology. Their inquiry began with the realization that iridium – an element rare on earth but common in asteroids – appears in the geological record in concentrations 20 to 160 times background levels at the time the dinosaurs became extinct (Alvarez, Alvarez, Asaro & Michel, 1980). "Some of the greatest achievements in science come from work at the boundaries of disciplines."[18] Unlike the example of the butcher, the baker, and the brewer, however, it is not always clear beforehand which groups need to share information.

Even in those cases where the connections are obvious, the necessary interactions may fail to occur. If the returns to individuals from more specialized interaction do not align with the returns to societies, then persons acting out of pure self-interest will not internalize the spill over effects -- the externalities -- that would otherwise benefit society. Voluntary fragmentation might then produce direct economic costs. For example, the benefits of knowledge produced by the “intellectual cross-pollination” of having academic researchers interact with business practitioners might spill over into government policy — a positive externality. These two groups will interact too infrequently, however, if they do not internalize the benefits they provide to other groups. In such situations, reduced specialization would be optimal from a purely economic perspective.[19]

Independent of the potential economic costs or benefits of integration, members of a society may wish to increase integration simply to maintain a degree of societal cohesiveness. With the customized access and search capabilities of IT, individuals can focus their attention on career interests, music and entertainment that already match their defined profiles, and they can read only news and analysis that align with their preferences.[20] Individuals empowered to screen out material that does not conform to their existing preferences may form virtual cliques, insulate themselves from opposing points of view, and reinforce their biases. Authors of collaborative

---

[18](Angier, 1985, citing Robert Hazen)

[19]This assumes that there is a “missing market” for cross-pollinating information since a third community could presumably subsidize any interaction which provided measurable benefits. One information paradox, however, is that a potential buyer cannot accurately assess the value of shared information without inspecting it, but having inspected it, the buyer cannot in good faith return it to the seller and claim also to know nothing of what he has seen(Arrow, 1962). It therefore seems probable that both positive externalities and missing markets exist which could lead to excessive balkanization and inefficiency.

[20]The Firefly website at http://www.ffly.com provides a demonstration of several powerful filtering and searching technologies, including a collaborative filtering tool which recommends music based on a user's past preferences and those of his or her "neighbors" in music-space.

filtering technology have long recognized its ability to both foster tribalism as well as a global village (Resnick, Iacovou, Suchak, Bergstrom & Riedl, 1994).

Internet users can seek out interactions with like-minded individuals who have similar values, and thus become less likely to trust important decisions to people whose values differ from their own. This voluntary balkanization and the loss of shared experiences and values may be harmful to the structure of democratic societies as well as decentralized organizations.

In addition, the stratification that results from stratification along a quality dimension raises a concern for equality. The gap between the information rich and the information poor can widen with virtual communities (Alstyne & Brynjolfsson, 1995). If diversity of interaction or equality of resources represent goals we would ascribe to our social planners, we need to consider what level of integration we deem most suitable for balancing our private interests as individuals and our shared interests as members of a community. Fragmentation in one or more dimensions of our interactions may or may not be desirable, but once achieved, it can be difficult to reverse. In any event, at this early stage of developing information infrastructure, no single scenario is inevitable. We can, and should, explicitly consider what we value as we shape the nature of our networks and infrastructure — with no illusions that a greater sense of community will inexorably result.

IX Appendix A – Glossary of Variables

Var.	Unit	Interpretation
$\mathcal{A}$	$\mathbb{Z}$	Access: The total number of individuals reachable in the population independent of capacity $\mathcal{C}$.
$\mu_{\mathcal{A}}$	$\mathbb{R}[0,1]$	Balkanized Affiliation: Fragmentation of group membership.
μ_C	$\mathbb{R}[0,1]$	Balkanized Communication: Fragmentation of paths to information.
μ_I	$\mathbb{R}[0,1]$	Balkanized Information: Fragmentation of knowledge resources.
$\mathcal{C}$	$\mathbb{Z}$	Channels: A proxy for bounded rationality indicating the total number of possible simultaneous connections.
i,j,k	$\mathbb{Z}$	Individuals: Members from the population whose size is $\mathcal{N}$.
k_{it}	$\mathbb{R}$	Knowledge: An information resource owned by individual i of type t.
k_t	$\mathbb{R}$	Knowledge: The total information of a given type or the sum of individuals' information along this dimension or $\sum_{i=1}^{\mathcal{N}} k_{it}$
$\mathcal{K}$	$\mathbb{R}^{\mathcal{T}}$	Knowledge: The total information available to a population, given by $\mathcal{K} = [k_1, k_2, ..., k_T]$.
$\mathcal{L}_{ij}$	$\mathbb{Z}$	Links: The shortest communications distance between i and j.
$\mathcal{M}(t)$	$\mathbb{Z}$	Members: The set of all individuals with access to type t information.
$\mathcal{N}$	$\mathbb{Z}$	Population size.
$\mathcal{P}_i$	$\mathbb{R}^{\mathcal{T}}$	Profile: A vector signifying the amount of information of each type available to individual i. In isolation, this is $\mathcal{P}_i = [0, 0, ... k_{it}, ... 0]$.
$\mathcal{S}_{ij}$	$\mathbb{R}[0,1]$	Similarity: The degree of overlap between information profiles, measured as $\text{Cos}(\Theta_{ij})$.
Θ_{ij}	$\mathbb{R}$	Theta: The angle between the information profiles of i and j in knowledge space.
$\mathcal{T}$	$\mathbb{Z}$	Types: The total number of information categories $[1, 2, ... \mathcal{T}]$

$\mathbb{R}$ - Real, $\mathbb{Z}$ - Integer.

X. Appendix B

A. Proofs

Proof 1: Without bounded rationality, $\mathcal{C} \geq \mathcal{N}$ so every agent can connect to every knowledge base and $\forall\ i, j$ we have that $\mathcal{P}i = \mathcal{P}j = \mathcal{K}$ thus $\|\ \mathcal{P}i - \mathcal{P}j\ \| = 0$ and $\mathrm{Cos}(\Theta_{ij}) = 1$. Also, if every agent has access to all topics, then $\forall\ t$, $\mathcal{M}(t) = \{1, 2, ...\ \mathcal{N}\}$ i.e. membership is the population. Therefore $\mu_A = 1 - (1/\mathcal{T})(1/\mathcal{T}-1)(\mathcal{T})(\mathcal{T}-1)(N^2/N^2) = 0$

Proof 2 -- To prove Proposition 2, we need to show that knowledge profiles for agents *i* and *j* in different communities diverge with rising access and that μ_A also rises. For geographically restricted access, we show $\mathrm{Cos}(\Theta_{ij}) \geq 0$ initially,[21] implying their knowledge profiles overlap. For geographically unrestricted access, we show $\mathrm{Cos}(\Theta_{ij}) = 0$ implying their profiles are completely orthogonal.

The number of connections of each type that an agent experiences is a hypergeometric probability distribution. Let *t* be the prevalence of a given type in a population of $\mathcal{N}$ when access provides $\mathcal{A}$ samples. Then the probability of *x* contacts to this type is given by:

2.a
$$\frac{\binom{t}{x}\binom{\mathcal{N}-t}{\mathcal{A}-x}}{\binom{\mathcal{N}}{\mathcal{A}}}$$

The mean of this distribution is $\mathcal{A}\,(t/\mathcal{N})$. For geographically unrestricted access, $\mathcal{A}(t/\mathcal{N}) \geq \mathcal{C}$ so that agents can use all channels to contact their preferred type. Including i's own resources, the knowledge profile for agent *i* is $\mathcal{P}_i = <0, 0, ...\ (\mathcal{C}+1)k_{it}, 0, ...\ 0>$ and similarly for *j* with $(\mathcal{C}+1)k_{jt}$ occuring in a different slot. The dot product of these vectors must be zero so $\mathrm{Cos}(\Theta_{ij}) = 0$ after unrestricted access.

For geographically restricted access we may assume that the mean of the distribution is less than capacity or $\mathcal{A}(t/\mathcal{N}) < \mathcal{C}$. Including an agent's own resources, there will be $\mathcal{A}(t/\mathcal{N})+1$ expected contacts of the preferred type, and $\mathcal{C} - \mathcal{A}(t/\mathcal{N})$ contacts spread over $\mathcal{T}$-1 indifferent types. For simplicity, if we assume that the k_{it} are all equal to a constant κ, then the initial knowledge profile is:

2.b
$$\mathcal{P}_i = \left\langle \frac{[C-\mathcal{A}\frac{t}{N}]}{(\mathcal{T}-1)}\kappa, \frac{[C-\mathcal{A}\frac{t}{N}]}{(\mathcal{T}-1)}\kappa, ...[\mathcal{A}\tfrac{t}{N}+1]\kappa, \frac{[C-\mathcal{A}\frac{t}{N}]}{(\mathcal{T}-1)}\kappa, ... \right\rangle$$

Again, the knowledge profile $\mathcal{P}j$ is similar with the contacts for the preferred type occuring in a different slot. The numerator of $\mathrm{Cos}(\Theta_{ij})$ is given by $\mathcal{T}$-2 indifferent terms and 2 preferred terms or:

[21] Only if geographic communities are completely balkanized initially will $\mathrm{Cos}(\Theta_{ij}) = 0$. Any level of prior integration during thitial conditions implies the inequality is strict.

2.c
$$\mathcal{P}_i \bullet \mathcal{P}_j = \frac{(\mathcal{T}-2)(\mathcal{C}-\mathcal{A}\frac{t}{N})^2\kappa^2}{(\mathcal{T}-1)^2} + \frac{2(\mathcal{C}-\mathcal{A}\frac{t}{N})(\mathcal{A}\frac{t}{N}+1)\kappa^2}{(\mathcal{T}-1)}$$

Using the same vectors, the denominator of $\mathrm{Cos}(\Theta_{ij})$ has $\mathcal{T}$-1 indifferent terms and only 1 preferred term or:

2.d
$$\|\mathcal{P}_i\|\|\mathcal{P}_j\| = \frac{(\mathcal{T}-1)(\mathcal{C}-\mathcal{A}\frac{t}{N})^2\kappa^2}{(\mathcal{T}-1)^2} + \frac{(\mathcal{A}\frac{t}{N}+1)^2\kappa^2}{1}$$

After simplification, the $\boldsymbol{\kappa}$s cancel and the ratio becomes:

2.e
$$\frac{\mathcal{P}_i \bullet \mathcal{P}_j}{\|\mathcal{P}_i\|\|\mathcal{P}_j\|} = \frac{(\mathcal{T}-2)(\mathcal{C}-\mathcal{A}\frac{t}{N})^2 + 2(\mathcal{T}-1)(\mathcal{C}-\mathcal{A}\frac{t}{N})(\mathcal{A}\frac{t}{N}+1)}{(\mathcal{T}-1)(\mathcal{C}-\mathcal{A}\frac{t}{N})^2 + (\mathcal{T}-1)^2(\mathcal{A}\frac{t}{N}+1)^2}$$

Note that the number of types is large so $\mathcal{T}$-2 > 0 and that due to restricted access the expected number of contacts $\mathcal{A}(t/\mathcal{N}) < \mathcal{C}$ so $[\mathcal{C}-\mathcal{A}(t/\mathcal{N})] > 0$. Thus all terms are non-negative, establishing the result that $\mathrm{Cos}(\Theta_{ij}) \geq 0$ and knowledge profiles overlap under restricted access.

For μ_A it is easier to consider the probability that two agents with different interests will join each other's communities. As access rises, this probability falls. The initial probability that *i* does not contact any member of j's community is 1 - $[\mathcal{C}-\mathcal{A}(t/\mathcal{N})]/(\mathcal{T}-1)$ and similarly for *j*. Thus, for sufficiently large populations, the probability that *j* and *i* are in different communities is

2.f
$$\left[1-\frac{(\mathcal{C}-\mathcal{A}\frac{t}{N})}{(\mathcal{T}-1)}\right]^2$$

As access $\mathcal{A}$ rises, the expected number of preferred contacts $\mathcal{A}(t/\mathcal{N})$ eventually exceeds $\mathcal{C}$ so that the probability *i* and *j* join different communities goes to 1. The balkanization index μ_A must therefore rise with increasing access, completing the proof.

The limit behavior of equation **2.e** provides useful intuition. Taking the limit as $\mathcal{C}\to\infty$ gives $\mathrm{Cos}(\Theta_{ij}) \to (\mathcal{T}-2)/(\mathcal{T}-1)$ which goes to 1 since $\mathcal{T} > \mathcal{C}$. This implies that infinite capacity gives everyone the same knowledge profile and provides indirect confirmation of Proposition 1. If, instead, we first take $\mathcal{T}\to\infty$ in equation **2.e** and hold $\mathcal{C}$ constant, then $\mathrm{Cos}(\Theta_{ij}) \to 0$, which implies that a very large number of types causes the expected profiles to diverge.

Proof 3: From Proposition 2, we know that under unrestricted (or "global") access the knowledge profile of an agent *i* with preferred contacts is $\mathcal{P}_i = [0, 0, \ldots \mathrm{C}\boldsymbol{\kappa}_t, \ldots 0]$. Let the new number of channels C' be given by $\mathcal{C}+\Delta$. Then if agents use the new channels to connect to additional knowledge bases of the same type and deepen their specialty the new knowledge profile becomes $\mathcal{P}_i = [0, 0, \ldots (\mathcal{C}+\Delta)\boldsymbol{\kappa}_t, \ldots 0]$ and similarly for $\mathcal{P}_j = [0, \ldots (\mathcal{C}+\Delta)\boldsymbol{\kappa}_s, \ldots 0, 0]$ with $s \neq t$. But then $\mathrm{Cos}(\Theta_{ij}') = \mathrm{Cos}(\Theta_{ij}) = 0$ and since no community has added new members $\mu_A' = \mu_A$ indicating that the respective communities are as just as balkanized as before.

Proof 4 -- We need to show that allocating more channels to one type causes knowledge profiles to diverge in different communities. In this case, let access be unrestricted so that only preferences matter. By the same logic as in Proposition 2, if t is the prevalence of a given type in a population of $\mathcal{N}$ with $\mathcal{C}$ samples, then the mean is $\mathcal{C}$ ($t/\mathcal{N}$). If agents are indifferent to their connections, this is true for any $\mathcal{A}$ and $\mathcal{C}$ (if access binds instead of capacity, simply replace $\mathcal{C}$ with $\mathcal{A}$ below). We can simplify our equations by assuming that the $\boldsymbol{\kappa}_{\mathrm{it}}$ are each equal in magnitude to a constant $\boldsymbol{\kappa}$ and that each of the various types t are equally probable. Since there are a total of $\mathcal{T}$ types, the likelihood of drawing a type t is $(t/\mathcal{N}) = (t/\mathrm{Tt}) = (1/\mathcal{T})$. The expected number of contacts by type is thus $(\mathcal{C}/\mathcal{T})$. Since agents reach their own knowledge bases with certainty, the expected knowledge profile of an agent i is $\mathcal{P}_i = \langle(\mathcal{C}/\mathcal{T})\boldsymbol{\kappa}, (\mathcal{C}/\mathcal{T})\boldsymbol{\kappa}, \ldots (1+\mathcal{C}/\mathcal{T})\boldsymbol{\kappa}, \ldots (\mathcal{C}/\mathcal{T})\boldsymbol{\kappa}\rangle$. For i and j in different communities, we have $\mathrm{Cos}(\Theta_{ij}) =$

4.a
$$\frac{\mathcal{P}_i \bullet \mathcal{P}_j}{\|\mathcal{P}_i\|\|\mathcal{P}_j\|} = \frac{(\mathcal{T}-2)\left(\frac{\mathcal{C}}{\mathcal{T}}\right)^2 \kappa^2 + 2\left(\frac{\mathcal{C}}{\mathcal{T}}\right)\left(1+\frac{\mathcal{C}}{\mathcal{T}}\right)\kappa^2}{(\mathcal{T}-1)\left(\frac{\mathcal{C}}{\mathcal{T}}\right)^2 \kappa^2 + \left(1+\frac{\mathcal{C}}{\mathcal{T}}\right)^2 \kappa^2}$$

After algebraic simplification, the $\boldsymbol{\kappa}$'s cancel and this expression reduces to:

4.b
$$\frac{2\mathcal{C}+\mathcal{C}^2}{2\mathcal{C}+\mathcal{C}^2+\mathcal{T}}$$

where, again, $\mathcal{C}$ is the number of channels and $\mathcal{T}$ is the number of types. This is the measure of overlap if agents are equally happy mixing with the population at large. If, on the other hand, agents are not indifferent to their connections but prefer to allocate $\mathcal{X}$ of their channels to a specific type then the expression for random association becomes

4.c
$$\frac{2(\mathcal{C}-\mathcal{X})+(\mathcal{C}-\mathcal{X})^2}{2(\mathcal{C}-\mathcal{X})+(\mathcal{C}-\mathcal{X})^2+\mathcal{T}}$$

For all X, $1 \le \mathcal{X} \le \mathcal{C}$ this implies the overlap between profiles diminishes. Note that if agents allocate all their channels to preferred types, then $\mathcal{X}=\mathcal{C}$ which leads again to complete balkanization, i.e. no overlap in knowledge profiles among agents from different communities. This establishes our contention that the stronger are an agent's preferences the greater is the degree of balkanization.

We can actually take the results a step further. Using the results from **2.e**, we can show the unusual result that severely restricting access has the same effect as imposing indifference. For severely restricted access, $\mathcal{A} < \mathcal{C}$ so that $\mathcal{A}$ binds in **2.e**. If we substitute $\mathcal{A}$ for $\mathcal{C}$ and use the same simplification to replace all instances of $t/\mathcal{N}$ with $\mathcal{T}$, then **2.e** reduces to:

4.d
$$\frac{2\mathcal{A}+\mathcal{A}^2}{2\mathcal{A}+\mathcal{A}^2+\mathcal{T}}$$

which resembles the equation for indifferent connections **4.b** with the caveat that all agents are below their capacity. Thus more restricted access can induce more diverse interaction. Note also that neither the results of Propostion 2 nor those of 4 depend on a homogeneous population distribution; non-uniform clustering gives similar results. In this example, setting $\mathcal{T}=t/\mathcal{N}$ makes the equations more tractable, but it only needs to be the case that some capacity is used to contact different types under restricted access for balkanization to rise with strong preferences under increased access. It is not necessary that types be uniformly distributed.

Proof 5 -- For specialization due to preferences, the proofs are identical to the proofs of Propositions 2 and 4 with $\mathcal{T}$ interpreted as grades in quality. For specialization due to veto power on the part of a destination community a detailed proof is provided in the full version of (Alstyne & Brynjolfsson, 1995). Briefly, they argue that if agents' opportunities for association are ranked along a single dimension, associations among relative peer groups constitute a Nash equilibrium. Agents in the top tier first commit to pair with one another. Then, having exhausted their options for contact, they become inaccessible to the next tier. The next tier becomes the most attractive set of partners and the exclusion process cascades. No agent can do better by altering their choices. Quality constitutes the single dimension considered for Proposition 5.

B. A Note on Measures

The measures can, in fact, move somewhat independently. Consider a completed table for Figure 5.

Index	Figure 5.A	Figure 5.B	Figure 5.C
Average ($\mathcal{S}_{ij}$)	0.69	0.44	0.25
μ_A	0.51	0.16	0
μ_C	0.72	0.72	0.22
μ_I	0.09	0.13	0.19

The table shows affiliation (μ_A) and communication (μ_C) become more balkanized but information resources (μ_I) become more integrated. Under random association in 5.A, the information resources are more decentralized and thus more fragmented than they are when collected into groups in 5.C so μ_I increases. Note, however, that groups no longer overlap; there is no inter-group affiliation so μ_A falls. Also, in the shift from 5.A to 5.B, the order of connections has changed but the total path length between agents has not so the communications index μ_C remains unchanged.

Another important observation is the tradeoff between complete range and sensitivity. A measure cannot have both full range (i.e. take on all values in $0 \le \mu \le 1$) and also be sensitive to population size. A population of 12 agents, for example, might be considered completely integrated when all 12 are connected and so an integration index might obtain its highest value, here 1 (or balkanization 0). But, there is a sense in which a fully integrated population of 120 is even more integrated than a fully integrated population of 12 (connections are increasing arithmetically). Thus one might prefer $\mu(12) > \mu(120)$ for two fully integrated communities. But then, the index needs to account for arbitrary size populations and so $0 \le \mu < 1$ for any finite population (i.e. its range is not closed and it cannot assume the value 1). As currently defined, certain of the indices do not have full range. They can be "tweaked" by changing certain coefficients and the range of summations to give the indices full range. For large populations, both formulations will be asymptotically the same. For small populations, they can differ substantially. When working with fixed populations we suggest choosing a measure with full range, but when working with varying populations we suggest choosing the more sensitive measure.

Bibliography

Alstyne, M. V. (1997). The State of Network Organization: A Survey in Three Frameworks. Journal of Organizational Computing, 7(3 (forthcoming)).
Alstyne, M. V., & Brynjolfsson, E. (1995, December 14-17). Communication Networks and the Rise of an Information Elite -- Do Computers Help the Rich get Richer? Paper presented at the International Conference on Information Systems, Amsterdam.
Alvarez, L. W., Alvarez, W., Asaro, F., & Michel, H. V. (1980). Extraterrestrial Cause for the Cretaceous-Tertiary Extinction -- Experimental Results and Theoretical Interpretation. Science, 208(4448), 1095-1108.
Angier, N. (1985, May 6). Did Comets Kill the Dinosaurs? Time.
Arrow, K. (1962). Economic Welfare and the Allocation of Resources for Invention. In R. Nelson (Ed.), The Rate and Direction of Inventive Activity, (pp. 609-625). New Jersey: Princeton University Press.
Baker, W. E. (1993). The Network Organization in Theory and Practice. In N. Nohria & R. G. Eccles (Eds.), Networks and Organizations, (pp. 397-429). Boston, MA: Harvard Business School Press.
Black, F., & Scholes, M. (1973). The Pricing of Options and Corporate Liabilities. Journal of Political Economy, 81(3), 637-654.
Burt, R. (1993). The Social Structure of Competition. In N. Nohria & R. G. Eccles (Eds.), Networks and Organizations, (pp. 57-91). Boston, MA: Harvard Business School Press.
Chao, J. (1995, June 7). Internet Pioneers Abandon World They Created. Wall Street Journal, pp. B8.
Donath, J. (1995, Nov. 5-9). Visual Who: Animating the Affinities and Activities of an Electronic Community. Paper presented at the ACM Multimedia, San Francisco, CA.
Freeman, L. C. (1979). Centrality in Social Networks: Conceptual Clarification. Social Networks, 1, 215-239.
Gaspar, J., & Glaeser, E. L. (1996). Information Technology and the Future of Cities : NBER.
Granovetter, M. (1973). The Strength of Weak Ties. American Journal of Sociology, 78(6), 1360-1380.
Hesse, B., Sproull, L., Kiesler, S., & Walsh, J. (1993). Returns to Science: Computer Networks in Oceanography. Communications of the ACM, 36(8), 90-101.
Hiltz, S. R., & Wellman, B. (1997). Asynchronous Learning Networks as a Virtual Classroom. Communications of the ACM, 40(9), 44-49.
Kaufer, D. S., & Carley, K. M. (1993). Communication at a Distance: The Influence of Print on Sociocultural Organization and Change. Hillsdale, NJ: Lawrence Erlbaum Associates.
Kremer, M. (1993). The O-Ring Theory of Economic Development. Quarterly Journal of Economics, 108(3), 551-575.
Kuhn, T. (1970). The Structure of Scientific Revolutions. (2nd ed.). Chicago: University of Chicago Press.
Leslie, J. (1995,). Mail Bonding. Wired, 2, 42-48.
Malone, T. W., & Smith, S. A. (1988). Modeling the Performance of Organizational Structures. Operations Research, 36(3), 421-436.
McLuhan, M., & Powers, B. R. (1989). The Global Village -- Transformations in World Life and Media in the 21st Century. New York: Oxford University Press.
Moffat, A. S. (1993). New Meetings Tackle the Knowledge Conundrum. Science, 259, 1253-55.
Resnick, P., Iacovou, N., Suchak, M., Bergstrom, P., & Riedl, J. (1994,). GroupLens: An Open Architecture for Collaborative Filtering of Netnews. Paper presented at the Computer Supported Cooperative Work.
Roth, A. E. (1984). The Evolution of the labor Market for Medical Interns and Residents: A Case Study in Game Theory. Journal of Political Economy, 92(6), 991-1016.
Simon, H. A. (1957). Models of Man, Social and Rational. New York: John Wiley & Sons.

Sims, C. (1996, Mon., May 27). A Web Entree for Peruvians without PCs. New York Times, pp. 29, 31.

Sproull, L., & Kiesler, S. (1991). Connections -- New Ways of Working in the Networked Organization. Cambridge: MIT Press.

Turkle, S. (1995). Life on the Screen -- Identity in the Age of the Internet. New York: Simon & Schuster.

Wellman, B., & Gulia, M. (1997). Net Surfers Don't Ride Alone: Virtual Communities as Communities. In P. Kollock & M. Smith (Eds.), Communities in Cyberspace, (Vol. (in Press),). Berkeley: University of California.

White, H. C., Boorman, S. A., & Breiger, R. L. (1976). Social Structure from Multiple Networks I -- Blockmodels of Roles and Positions. American Journal of Sociology, 81(4), 730-780.

Wulf, W. A. (1993). The Collaboratory Opportunity. Science, 261(5123), 854-855.