ノード間関係の類似度を定量化するネットワーク間狭小性

森岡 淳† 古川 正志‡ 山本 雅人‡ 鈴木 育男‡
†‡北海道大学大学院情報科学研究科

1 はじめに

人々は様々な人と多様な形で関わり合い，社会を形成する．その関係をネットワークとして捉え，その構造の特性から人々のコミュニケーションを理解するための様々な研究が行われている．Watts らはそのようなネットワークから，スモールワールド性を発見した [1].

しかし，現実のコミュニケーションは同一の集団の中だけで行われているものではない．人々は幾多の集団に所属して生活している．例として，職場，家庭，プライベートでのコミュニケーションの形態は異なる．

このような実世界での関係は同一ノードを持つ複数のネットワークで表すことができる．これらの各ネットワークで同一ノード同士の距離関係が全く何の要因にもよらず完全に独立に決まるのであれば，ランダムに決まるのと同等である．しかし，何らかの性質によって距離関係が決まるならば，そのような性質の違いからネットワーク間の距離を測ることが可能である．

我々はこのような複数ネットワーク空間の異なるネットワーク間での同一ノード間の位置関係の類似度を定量化し，研究を進めてきた [3]. これをネットワーク間狭小性と名づけた．しかしながら，それを様々なネットワークに当てはめた結果，正しくノード間関係の類似度を測ることが難しい場合があることがわかった．本研究では，ネットワーク間狭小性をいくつかの実験を通して検証し，より妥当性のある式を用いて，ネットワーク間狭小性を再定義する．

2 ネットワーク間狭小性

2.1 同一ノードを持つネットワーク間の定義

本研究で用いる 2 つのネットワークを，1 つ以上のラベルを共有する 2 つのラベル付きネットワーク G_i，G_j として考える．ここで，V_i は G_i の頂点集合，E_i は G_i の辺集合である．ネットワーク $G_i = (V_i, E_i)$ のノードが持つラベルを $\psi_i(x)$ と書く．全ての $x \in V_i$，$y \in V_i$ について，$\psi_i(x) = \psi_j(y) \Leftrightarrow x = y$ とする．任意の 2 ネットワーク G_i，G_j について $\forall x \in V_i$，$\forall y \in V_j$ であるならば，$\psi_i(x) = \psi_j(y) \Leftrightarrow x \sim y$ とする．以降，このような x，y を同一ノードとする．

2.2 ネットワーク間狭小性の定量化

本研究で用いる複数ネットワーク空間には，各ネットワーク間に同一ノードの存在を前提とする．ネットワーク G_i でのノード x の近隣ノード y はネットワーク G_j ではノード x からみてどれくらい離れているかは明らかにされていない．これらを解明するために，本研究では同一ノードの個数だけではなく，同一ノード間の距離が両ネットワークで近ければ近いほどネットワーク間の距離が狭いと定義する．また，ネットワーク G_i においてノード x から距離が大きく離れている同一ノードがネットワーク G_j で同じ距離にいる場合よりも，ネットワーク G_i においてノード x の一次近傍のノードがネットワーク G_j で同じ距離である場合の方がネットワーク間の距離が狭いと定義する．この理由は，あるノードに対する影響力は，距離が近ければ近いほど高いからである．

そのために，まず $v(v \in G_i, G_j)$ の G_i，G_j におけるパス長 k 内の近傍，$\Gamma_i^k(v)$，$\Gamma_j^k(v)$ の内，同一ノードが含まれる割合 NS_{ij}^k を式 (1) で定義する．

$$NS_{ij}^k(v) = \frac{\left|\Gamma_i^k(v) \cap \Gamma_j^k(v)\right|}{f\left(\Gamma_i^k(v), \Gamma_j^k(v)\right)} g(k) \tag{1}$$

$\Gamma_i^k(v)$ は v_l のネットワーク G_i での k 次近傍内のノードの集合である．$f\left(\Gamma_i^k(v), \Gamma_j^k(v)\right)$ は両ネットワークでのパス長 k 内の近傍ノード数の関数．$g(k)$ は k の関数で近傍を広げるにつれて，ネットワーク狭小性の増分を抑えるために用いる．これらの関数によってネットワーク間狭小性が大きく左右される．

NS_{ij}^k を用いて，ネットワーク間狭小性を定量化する．ネットワーク G_i，ネットワーク G_j 間の同一ノード v のネットワーク間狭小性 $NS_{ij}(v)$ を式 (2) に示す．

$$NS_{ij}(v) = \frac{1}{\sum_{k=1}^{R} g(k)^{-1}} \sum_{k=1}^{R} NS_{ij}^k(v) \tag{2}$$

ここで $R = \max\{\mathrm{rad}(v|i), \mathrm{rad}(v|j)\}$ であり $\mathrm{rad}(v|i)$ は i におけるノード v の半径である．$k = R$ のときは，v が G_i，G_j で到達できる全てのノードが $\Gamma_i^R(v)$，$\Gamma_j^R(v)$ に含まれる．$\sum_{k=1}^{R} g(k)^{-1}$ は，$NS_{ij}(v_l)$ の最大値を 1 とするように

Network Smallworldness quantifying relations similarity between nodes
†Jun Morioka ‡Masashi Furukawa ‡Masahito Yamamoto ‡Ikuo Suzuki
†‡Graduate School of Infomation Science And Technology, Hokkaido University

正規化するため用いた．これを両ネットワークに含まれる全ての同一ノードについて計算することによってネットワーク G_i，ネットワーク G_j 全体のネットワーク間狭小性 NS_{ij} を式 (3) で求める．

$$NS_{ij} = \frac{1}{n_{ij}} \sum_{l=1}^{n_{ij}} NS_{ij}(v_l) \tag{3}$$

式中の n_{ij} はネットワーク G_i，ネットワーク G_j にともに存在するノードの数である．式 (3) は，式 (2) の全同一ノード平均の算出に相当する．以後，この値をネットワーク間狭小性とし，NS と略す．

3 新しい式の提案

上で述べた $f\left(\Gamma_i^k(v), \Gamma_j^k(v)\right)$ と $g(k)$ を具体的に定める．

まず，$f\left(\Gamma_i^k(v), \Gamma_j^k(v)\right)$ を取り上げる．従前の式は以下である [3]．

$$f\left(\Gamma_i^k(v), \Gamma_j^k(v)\right) = \frac{\left|\Gamma_i^k(v)\right| + \left|\Gamma_j^k(v)\right|}{2} \tag{4}$$

しかしながら，この式を用いると両ネットワークでの次数の違いが大きいと NS が小さくなってしまう問題がある．そこで，以下の式を新たに提案する．

$$f\left(\Gamma_i^k(v), \Gamma_j^k(v)\right) = \min\left(\left|\Gamma_i^k(v)\right|, \left|\Gamma_j^k(v)\right|\right) \tag{5}$$

この式は，近傍内のノード数が小さいほうのみを使用するため，次数の違いは影響しなくなる．

次に，$g(k)$ の満たすべき条件を考える．$g(k)$ は近傍が広がるにつれてネットワーク狭小性への影響を小さくするために用いている．よって，k に関する減少関数である必要がある．それを踏まえ，以前まで $g(k)$ には $\frac{1}{k}$ を用いていた [3]．そのほかの候補として $\frac{1}{k^2}$，$\frac{1}{2^{k-1}}$，$\frac{2}{k+1}$ があげられる．

4 検証実験

3 章で提案した式の妥当性を検証するために実験を行う．なお，各実験で用いたネットワークはノード数 200，リンク数 600 の ER ランダムモデル [2] で，各実験結果は 100 回試行し平均を算出したものである．$f\left(\Gamma_i^k(v), \Gamma_j^k(v)\right)$ について，式 (4) と式 (5) の違いを調べる実験を以下の手順で行う．最初に同一のネットワークを 2 つ用意する．これらの NS は最大の 1 である．そして，2 つの実験を行う．1 つ目の実験では一方のネットワークのみに，ランダムにノードを追加すると同時にそのノードとランダムに選んだ 3 つのノードとリンクをつなげる．100 ノードを追加するごとに，各式で NS を計算する．なお，$g(k) = \frac{1}{k}$ とした．図 1 中の Exp.1 に示した結果を見ると，式 (4) がノードが増えるにしたがって指数関数的に減少するのに比べて式 (5) は常に 1 である．これより，次数の違いが影響しないことが示せた．2 つ目の実験は，両方のネットワークに同一ノードではないノードとリンクを追加する．その他の条件は 1 つ目の実験と同じである．結果は図 1 中の Exp.2 に示した．どちらの式も若干の違いはあるが指数関数的に減少している．これらの実験より式 (5) は次数の違いによる影響のみを取り除く式であるとする．

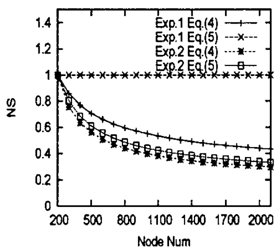

図 1: ノード数と NS の変化

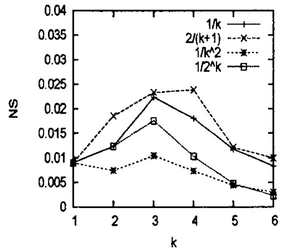

図 2: k と NS の変化

次に，$g(k)$ の各式の違いを調査する．両ネットワークの 20 % をランダムに同一ノードとし，k が 1 ずつ増えることによる NS の増分を 4 つの式を用いて計算する．なお，式 (5) を用いた．結果を図 2 に示す．これを見ると，各々の式の特徴が表れている．まず，$\frac{k+1}{2}$ のみ $k = 4$ のときにピークが現れる．他の 3 つの式が $k = 3$ のときにピークが現れるので，$\frac{k+1}{2}$ はより遠くのノードを見たときにピークが現れることがわかる．また，$\frac{1}{2^{k-1}}$ は全体的に $\frac{1}{k}$ より小さい値をとり，$\frac{1}{k^2}$ はそれよりも小さい値をとる．しかし，$k = 6$ になると $\frac{1}{2^{k-1}}$ の方が小さくなっているので，より大きい k では $\frac{1}{2^{k-1}}$ は最も低い値をとることがわかる．

5 おわりに

ノード間関係の類似度を定量化するネットワーク間狭小性を再定義し，そのパラメーターによる違いを実験により示した．$f\left(\Gamma_i^k(v), \Gamma_j^k(v)\right)$ については，よりノード間関係の類似度に特化した式 (5) を用いるのが妥当である．また，$g(k)$ についてはどの式が妥当かを決めず，各々の目的に合わせて選択するのがよい．

参考文献

[1] D. J. Watts, Small Worlds: The Dynamics of Networks Between Order and Randomness, Princeton Univ. Press, Princeton, NJ, (1999).

[2] P.Erdös, A.Rényi, publ,Math(1959).

[3] 森岡淳，吉井伸一郎，古川正志，ネットワーク間狭小性の定量化によるタグの間接共起解析，北海道情報処理シンポジウム 2007(2007)