

# 貪欲到達中心性によるネットワーク探索性能の特徴付け Characterizing Network Search Performance by Greedy Reachability Centrality

宋 鵬<sup>†</sup>  
Peng Song

齊藤 和巳<sup>†</sup>  
Kazumi Saito

池田 哲夫<sup>†</sup>  
Tetsuo Ikeda

青山 一生<sup>‡</sup>  
Kazuo Aoyama

## 1. はじめに

多様な現実ネットワークの共通性質として、六次の隔たり (Six Degrees of Separation) の性質に代表されるスモールワールド性 [1] が注目されている。特にその機能として重要なのは、ネットワーク上でターゲットの位置が明確に分からなくても、任意のノードから近傍情報のみを利用して、ほぼ最短に近いパスを見つけ出し、メッセージ配信が可能なことである。しかし、Watts-Strogatz [2] や、Kleinberg [3] などネットワーク構成法に関する研究は進展したものの、探索の観点でのネットワーク構造分析に関する研究は殆ど知られていない。

本稿では、工学的な視点に基づき、メッセージ配信をオブジェクトの類似探索問題と捉え、引込領域中心性 (basin centrality) と貪欲到達中心性 (greedy reachability centrality)、及び、それぞれの累積分布でネットワーク構造を分析する方法を提案する。具体的には、画像を構成する 4 つの MPEG-7 ビジュアル属性のデータ CoPhIR [4] を利用し、オブジェクト毎に  $k$  個の最類似オブジェクトを結合させて構成する  $k$ -NN ( $k$ -Nearest Neighbour) ネットワークを対象に、提案法の有効性を評価する。

## 2. 分析方法

探索対象のオブジェクト集合を  $\mathcal{V}$  とし、オブジェクトペア  $u, v \in \mathcal{V}$  に対し類似度  $s(u, v)$  が定義されているとする。また、被探索ネットワークを、類似度に基づきオブジェクトを結合する辺の集合  $\mathcal{E}$  を用いて  $G = (\mathcal{V}, \mathcal{E})$  とし、各オブジェクトを対応するノードと同一視する。一方、ノード  $u \in \mathcal{V}$  の  $G$  での隣接ノード集合を  $\mathcal{A}(u)$  で表し、与えられたクエリ  $q$  に対し、ノード  $u$  と  $\mathcal{A}(u)$  の中で、 $q$  との最類似するノードを  $g(u; q)$  とし次式で定義する。

$$g(u; q) = \arg \max_{v \in \{u\} \cup \mathcal{A}(u)} \{s(v, q)\}. \quad (1)$$

また、正の整数  $h$  に対し、 $g^{(h)}(u; q) = g(g^{(h-1)}(u; q); q)$  を定義する。ただし、 $g^{(0)}(u; q) = u$  とする。このとき、任意の探索対象オブジェクト  $v \in \mathcal{V}$  をクエリとし、オペレータ  $g(\cdot; v)$  を再帰的に適用し、 $v$  へ到達する (探索可能、または配信可能) ノード集合  $\mathcal{BS}(v)$  を次式で定義する。

$$\mathcal{BS}(v) = \{u \in \mathcal{V} \mid \exists h, v = g^{(h)}(u; v)\} \quad (2)$$

以降では、 $\mathcal{BS}(v)$  の要素ノード数を  $bsc(v) = |\mathcal{BS}(v)|$  とし、ノード  $v$  の引込領域中心性 (basin centrality) と呼ぶ。

次に、探索開始ノード  $u \in \mathcal{V}$  に着目し、 $u$  から到達可能なクエリ集合  $\mathcal{GR}(u)$  を次式で定義する。

$$\mathcal{GR}(u) = \{v \in \mathcal{V} \mid u \in \mathcal{BS}(v)\}. \quad (3)$$

同様に、 $\mathcal{GR}(u)$  の要素ノード数を  $grc(u) = |\mathcal{GR}(u)|$  とし、ノード  $u$  の貪欲到達中心性 (greedy reachability centrality) と呼ぶ。また、引込領域中心性  $bsc(v)$  と貪欲到達中心性  $grc(u)$  に対し、中心性値 (ノード数  $n$ ) での累積分布  $C_{bs}(n)$  と  $C_{gr}(n)$  をそれぞれ次式で定義する。

$$C_{bs}(n) = |\{v \in \mathcal{V} \mid bsc(v) \geq n\}|, \quad (4)$$

$$C_{gr}(n) = |\{u \in \mathcal{V} \mid grc(u) \geq n\}|. \quad (5)$$

本稿では、累積分布  $C_{bs}(n)$  と  $C_{gr}(n)$  のそれぞれをプロットし分析する方法を提案する。

## 3. 実験による評価

実験データとして、写真画像のコンテンツデータベースである、CoPhIR(Content-based Photo Image Retrieval) [4] を利用した。全 105,998,779 画像の中から 10,000 画像をランダムに抽出し、この抽出画像に対し、4 つの MPEG-7 ビジュアル属性 ColorStructure (cs), ScalableColor (sc), EdgeHistogram (eh), HomogeneousTexture (ht), 及び、これらを混合 (mx) で記述された 5 種のデータを用いた。ここで、cs, sc は色に関する属性で、ともに 64 次元のベクトル、eh, ht は模様に関する属性で、順に 150 次元と 62 次元のベクトル、mx は前述のベクトルの重み付き結合で 340 次元のベクトルからなる。

本実験では、オブジェクト毎に  $k$  個の最類似オブジェクトを結合させて構成する  $k$ -NN ネットワークで提案法を評価した。図 1 では、各属性に対し、 $k$  の値を 3 から 9 まで変化させ、引込領域中心性  $bsc(v)$  の累積分布  $C_{bs}(n)$  を比較した。図より、どの属性においても、 $k$  の値を大きくすれば、引込領域中心性の高いノードは相対的に増加するが、その増加率は  $k$  の値が小さいときの方が顕著なことが分かる。また、例えば  $k = 9$  のとき、引込領域中心性値が 3,000 以上となるノード数は、属性 cs, eh, mx, sc では全体の半数程度であるのに対し、属性 ht では 9 割程度に達していることが分かる。すなわち、属性 ht では、相対的に引込領域中心性の高いノードが多いことが分かる。

図 2 では、図 1 と同様に、各属性に対し、 $k$  の値を 3 から 9 まで変化させ、貪欲到達中心性  $grc(v)$  の累積分布  $C_{gr}(n)$  を比較した。この図からも、どの属性においても、 $k$  の値を大きくすれば、貪欲到達中心性の高いノードは相対的に増加し、その増加率は  $k$  の値が小さいときの方が顕著なことが分かる。一方、引込領域

<sup>†</sup>静岡県立大学 経営情報学部

<sup>‡</sup>NTT コミュニケーション科学基礎研究所

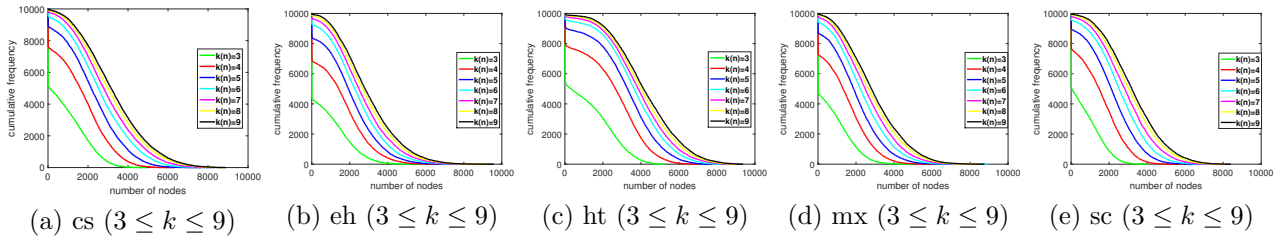
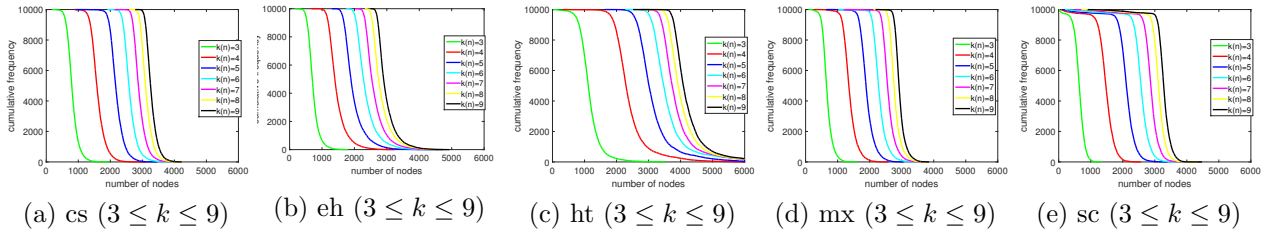
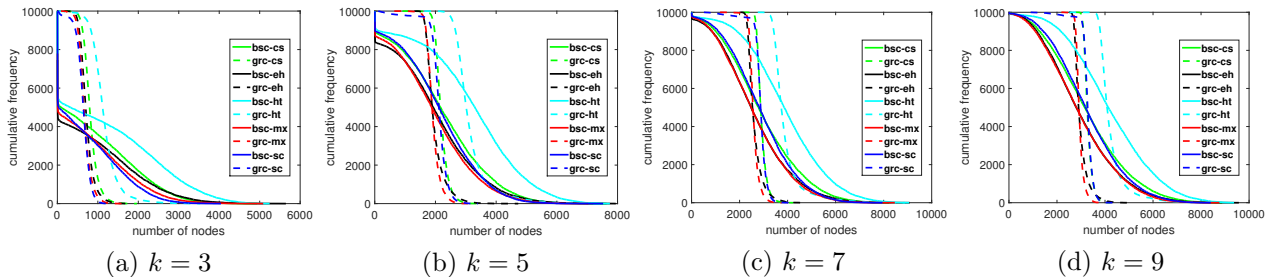
図1: 引込領域中心性  $bsc(v)$  の累積分布  $C_{bs}(n)$  の比較図2: 貪欲到達中心性  $grc(u)$  の累積分布  $C_{gr}(n)$  の比較

図3: 引込領域中心性と貪欲到達中心性によるデータ属性間での比較

中心性と貪欲到達中心性の累積分布を比較すれば、例えば  $k = 9$  のとき、どの属性においても引込領域中心性値は 6,000 程度以下で幅広く分布するのに対し、貪欲到達中心性値については、例えば cs では、3,000 から 3,500 程度の狭い範囲に分布していることが分かる。すなわち、引込領域中心性と貪欲到達中心性の分布は異なる性質を持つことが分かる。

図3には、 $k \in \{3, 5, 7, 9\}$  とし、引込領域中心性と貪欲到達中心性によるデータ属性間での比較を示す。明かに、どの  $k$  においても、属性 ht には、引込領域中心性でも貪欲到達中心性でも値の大きいノードが顕著に多いことが確認できる。すなわち、オブジェクト間の類似度構造は、これら中心性値の分布に影響を与えることが示唆される。

#### 4. おわりに

本研究では、オブジェクトの類似探索の観点でのネットワーク構造分析に向け、引込領域中心性と貪欲到達中心性、及び、それぞれの累積分布でネットワーク構造を分析する方法を提案した。5種の CoPhIR データ [4] を用いた  $k$ -NN ネットワークでの実験では、引込領域中心性と貪欲到達中心性の累積分布は、相対的に、前者は広く後者は狭いことともに、データ種に依存して、分布が異なることを確認した。今後は、さらに多

様なデータやネットワークを用いた評価実験を進める。

謝辞 本研究は、科学研究費補助金基盤研究(C)(No.17K00156)の助成を受けた。

#### 参考文献

- [1] Milgram, S., "The small world problem," *Psychology Today* 2, 60–67(1967).
- [2] Watts, D. J., Strogatz, S. H. "Collective dynamics of 'smallworld' networks," *Nature* 393, 440–442, 1998.
- [3] Kleinberg, J., "Complex networks and decentralized search algorithms," *Proc. Int'l. Congress of Mathematicians*, 2006.
- [4] Bolettieri, P. et al., "CoPhIR: a test collection for content-based image retrieval," arXiv: 0905.4627v2, 2009.