複雑ネットワークのフラクタル性と renormalization

6/29 社会ネットワーク勉強会

内田 誠 uchida@race.u-tokyo.ac.jp

Myself & my interest

- 内田 誠 (うちだ まこと) / 23歳 (修士2年次)
- 所属
  - 東京大学 工学系研究科 環境海洋工学専攻
  - 人工物工学研究センター デジタル価値工学部門 白山研究室
- 研究テーマ
  - 製品やサービス設計のためのユーザーモデルの構築について
    ……including ユーザー同士の実のネットワーク構造と、その中での相互作用による情報や知識の伝播現象とか、行為や意思決定の創発現象とかそのシミュレーション (MAS) とか。
  - 実のネットワーク構造を既知として、早いとこその上での情報伝播とか相互作用のモデルを立てたシミュレーションをしたい。
  - しかし、扱っているデータが大規模すぎてネットワーク構造全部を扱おうとすると死ぬ。なんとかして、実のネットワーク独自の有意な構造を残したまま相似縮小できる方法論がないものだろうか。
  - 複雑ネットワークのフラクタル性を利用した相似縮小(renormalization) ができないかな。

Self-similarity of complex networks

- Chaoming Song(1, Shlomo Havlin(2, Herman A. Makse(1
  - 1) Levich Institute and Physics Department, City College of New York.
  - 2) Minerva Center and Department of Physics, Bar-Ilan University.
- Nature vol. 433 (2005), pp.392-395.
- *…現実の多くのネットワークは'scale-free'の性質が見られる。しかし、ネットワークの直径に比べて頂点数が指数的に増える (scale-free ではない) という'small-world'の性質のため、複雑ネットワークでは不変性や自己相似性は持たないと信じられてきた。ところが、現実のネットワークを分析してみたところ、どの距離のスケールでも自己相似性を示すパターンがあることが明らかになった。….*

Two fundamental Properties of real complex network (1 of 2)

- Small-world Property
  - ネットワークの頂点数に比べて直径*が対数的に増加する性質。
    *直径: ネットワーク中でもっとも遠いノード同士の距離。

$$\bar{l} \approx \ln N, \qquad N \approx e^{\bar{l}/l_0}$$

$\bar{l}$: diamiter of the network

N : Number of vertices

  - 頂点数と直径の関係がそれらの規模によって異なり、自己相似性は示さない。

Two fundamental Properties of real complex network (2 of 2)

- Scale-free Property
  - ネットワーク内の頂点の次数の確率分布が冪乗則 (power-law distribution) に従う性質。
    (ちなみに累積確率分布も冪乗則になる)

$$p(k) \approx k^{-\gamma}, \qquad P(k \geq) \approx k^{-\gamma'}$$

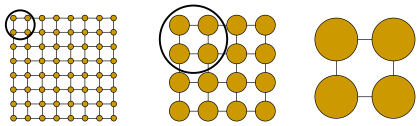
    - 現実の多くのネットワークでは2 < γ < 3くらい。
  - 次数と確率の関係はスケールによって不変で、自己相似性を示す。

距離と頂点数の関係はスケールフリーではなかったのに、なぜ次数分布はスケールフリーになり、自己相似性を示すのか？
⇒なにかスケール不変な性質を内在するのではないか？

Cluster and Fractal dimension. (1 of 3)

- ネットワークの中に存在するクラスタを考え、その自己相似性を観察する。クラスタを見つける方法は･･･
  - 'box-counting' method
  - 'cluster-growing' method

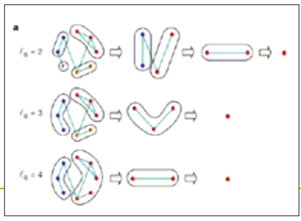
たとえば、2次元正方格子の場合は自己相似性を有する。



Cluster and Fractal dimension. (2 of 3)

- ‘box-counting’ method
  - ある’box-size’を決定し、boxの直径がbox-sizeを越えないようにネットワークをboxで区切る。
  - このとき、N_Bをboxの数、l_Bをbox-sizeとすると、フラクタル指数d_Bは以下の式で定義される。

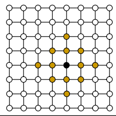
$$N_B \approx \ell_B^{-d_B}$$



Cluster and Fractal dimension. (3 of 3)

- ‘cluster-growing’ method
  - ランダムに頂点を決定し、その頂点から距離l以内にあるノードでクラスタを形成(ランダムに選んだ頂点のl次のego-centricネットワーク) 。
  - 充分多いクラスタを作り、それに含まれる平均の頂点数を$\langle M_c \rangle$とすると、フラクタル次元d_fは以下の式で定義される。

$$\langle M_c \rangle \approx \ell^{d_f}$$



2次元正方格子上でl=2の時

  - small-worldの性質を持つ場合はd_fは∞。
  - 複雑でない規則的なネットワーク (木とか格子とかランダムネットワークとか) では、d_Bとd_fは本質的に同じものになるが、複雑ネットワークでは意味合いが異なる。

Mean ‘mass’ of boxes

- ‘box-counting/covering’ method と’cluster-growing’ methodで作ったクラスタに含まれる頂点数の平均を比べる。
  - ‘box-counting/covering’ method

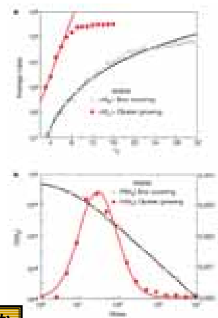
$$\langle M_B(\ell_B) \rangle = N/N_B(\ell_B) \approx \ell_B^{d_B}$$

↑冪乗的。確率分布に冪乗則 (scale-free性) が出る。

  - ‘cluster-growing’ method

$$\langle M_c(\ell_B) \rangle \approx e^{\ell_B/\ell_1}$$

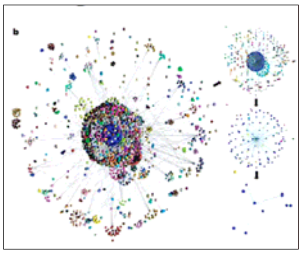
↑指数的。確率分布が対数正規分布になる。



‘box-counting’ methodでは全ての頂点が平等にどれかのboxに含まれるが、’cluster-growing’ methodではハブがboxに含まれやすくなるため。

‘box-counting’ method (1 of 4)

- WWWネットワークに’box-counting’ methodを適用した例。
  - N = 392340, l_B = 3.



Boxを1つのノード、box同士をつなぐリンクを新たなリンクとして考えれば、ネットワークの相似縮小(re-normalization)が可能か？

‘box-counting’ method (2 of 4)

- **実ネットワークサンプル**
  - WWW
  - Actors
  - Protein-protein interaction.
    - H. sapiens
    - E. coil
  - Cellular
    - A. fulgidus
    - E. coil
    - C. elegans
- **分析指標**
  - $N_B(l_B)$ / N と l_Bの関係。

$$N_B \approx \ell_B^{-d_B}$$

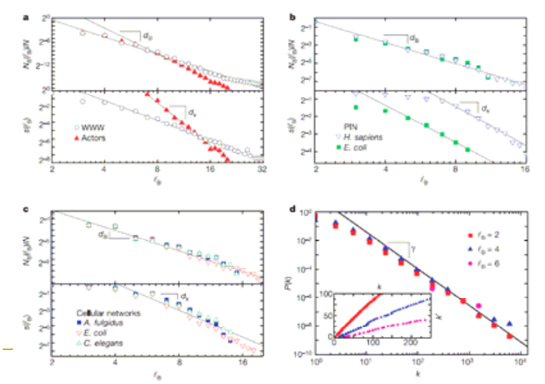
  - kとk’の関係。
    - k’はbox同士のリンクの次数、kをそのboxに含まれるノードの最大の次数。このとき、kとk’の比$s(l_B)$とl_Bの関係を調べる。

$$k \to k' = s(\ell_B)k \qquad s(\ell_B) \approx \ell_B^{-d_k}$$

    - k’の確率分布

$$P(k) \to P(k') \approx (k')^{-\gamma}$$

‘box-counting’ method (3 of 4)



'box-counting' method (4 of 4)

- $N_B(l_B)$ / N と l_Bの関係
  (フラクタル次元d_B)。
  - WWW d_B＝4.1
  - Actor d_B＝6.3
  - E.coli PIN d_B＝2.3
  - H. Sapiens PIN d_B＝2.3
  - Cellular Network d_B＝3.5
- kとk'の関係
  - WWW d_k＝2.5
  - Actor d_k＝5.3
  - E. coli PIN d_k＝2.1
  - H. Sapiens PIN d_k＝2.2
  - Cellular Network d_k＝3.2

d_Bは自己相似の構造の複雑さを表わす。自己相似構造を示すクラスタが大きければ、それだけ残るリンクは少なくなり、d_kは大きくなる 。ところが、 l_Bを変化させてクラスタの粒度を変えても、次数分布の冪乗則は崩れていないところに注目！！

Renormalization by 'box-counting/covering' method

- Renormalization前に次数kを持つノード数をn(k)、boxをノードに読み替えたrenormalization後に次数k'を持つノード数をn(k')とすると、P(k)とP(k')が同じ冪乗則に従っていたことより、 $n(k)dk = n'(k')dk'$
- 両辺を積分して、k'=skを代入すると、 $n(k) = s^{1-\gamma} n'(k)$
- ということは、renormalizingによってNあったノードがN'になる。 $N' = s^{\gamma-1} N$
- さらに、sはl_B の関数であったから、結局 $N' = N_B(\ell_B)$
- これより、γについて解くと、
  $$\gamma = 1 + d_B/d_k$$
- ⇒複雑ネットワークの冪指数γが、スケール非依存の自己相似性を表わす新たな指標d_Bとd_kによって表現・説明できる可能性がある！！

Conclusion

- 'small-world'と'scale-free'の両方の性質を持つ現実の複雑ネットワークも、length-scaleの自己相似性を持つ。
- 'box-counting' methodによって複雑ネットワーク中のノードをクラスタリングすると、次数のスケールだけでなく距離のスケールでも自己相似性が表れる。そのクラスタを新たなノードとすることで、ネットワークの相似縮小(renormalization)が可能になる。
- 'box-counting' method によるrenormalizationの結果、距離のスケールの自己相似性だけでなく、次数のスケールフリー性も保たれる
- 'scale-free'の性質の元になっていた次数の冪分布の冪指数γは、実は距離のスケールの自己相似性に関する別のより基礎的なパラメータで説明できる可能性がある。