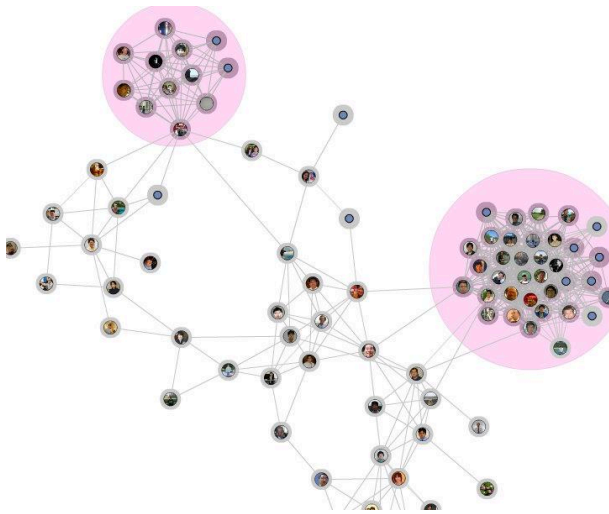


複雑ネットワークの構造を明らかにする統計科学

下平英寿

1. 複雑ネットワークって何？

私のFacebookアカウントには現在169人の友達が登録されている。みな現実世界の友達だ。169人の友達関係を頂点と枝からなるグラフ(ネットワーク)で表現する。



各枝の両端の二人は互いに友達である。ランダムに枝がついているわけではなく、何らかの構造がある。二つの大きなクラスタ(頂点が密に結合した集合)は、研究室の学生グループと、高校の同窓生グループであるが、その他にも小さなクラスタがいくつか見られる。このような現実のネットワークにはハブ(多数の枝を持つ頂点)があって、枝数 k の分布は裾の重い「べき乗則」 $\propto k^{-\gamma}$ に従う。ハブを通して意外と多くの人々が友達の友達だったりする。頂点数 n のときすべての頂点が互いに最短距離 $\propto \log n$ の枝数でつながる狭い世界だ。

2. 大規模ネットワークの可視化

ユーザー総数8億人を超えるFacebook全体を先ほどと同じようにグラフにするとどうなるだろうか？ウェブ全体はどうなるか？このような巨大なネットワークの構造を我々はどうかしたら把握できるだろうか？実際にグラフを描くと8億どころか1万人程度のネットワークでさえゴチャゴチャにからまった毛玉のようになってしまう。大規模な複雑ネットワークの構造を見えるようにする研究はまだ十分

に発展しておらず大きなチャレンジである。

3. ネットワーク構造を反映するページランク

ウェブページの重要度を測るGoogleのページランクは、ランダムにネットサーフィンしたときに各ページを見る確率に相当する。これはウェブ全体のグラフを巨大な行列で表現したときの固有ベクトルとして計算される。あのGoogleの中核技術は線形代数の基礎知識で簡単に理解できるし、君らにだって新しいアイデアで成功するチャンスは大いにある。

4. ネットワークの統計推測

リンク(グラフの枝)の多い人気サイトはさらにリンクを増やしてハブとなる。新しいリンクを獲得する確率がリンク数に比例すると仮定すれば、枝数がべき乗則に従う複雑ネットワークになる。下平研では、このようなグラフの生成メカニズムを推測する統計手法を提案している。グラフのクラスタが本当にコミュニティなのかそれとも見せかけなのかを測る信頼度の研究も行っている。

5. 未知のネットワーク構造を推測する

ヒトゲノムの全塩基配列はたった3G文字(6Gビット)なので、デジカメの1GBのメモリに余裕で保存できる。この読み取りが完了して久しいが、細胞内で遺伝子が互いの活動を制御するネットワークはまだよくわかっていない。グラフ構造の統計推測では従来ランダムグラフを仮定していたが、下平研ではより現実に近い複雑ネットワークを事前分布と仮定するベイズ推測を提案している。

6. データの要素数が負のシミュレーション!?

様々な生物のゲノムを比較して進化の系統樹(これもグラフ)を推測する分野では、下平研の提案した統計的信頼度が世界標準になっていて、論文が数千回引用されるほど利用されている。確率的なシミュレーションで生成するデータの要素数を観測データの要素数 $\times(-1)$ にするだけで精度が向上する。負の要素数なんてありえないところがこの理論のスゴイところだと自負している。アルゴリズムは実装できるし、確率分布の空間の距離とか曲率など幾何的な議論によって数学的に証明されている。