

4.2 部グラフの構築実験

実際のスレッドを用いて、提案手法の評価をスレッド名を用いた場合と本文を用いた場合でそれぞれ行う。

4.1 実験の設定

実験に使用するスレッドとして、2ちゃんねる[4]のyoutube板より表1に示すスレッドを取得する。取得するスレッドの条件としてスレッド名に”youtube”，”ニコニコ動画”，”ニコニコ生放送”のいずれかが含まれるものを選択した。これらのグループはいずれもキーワードノードを介して同グループ内全てのスレッドノードの組み合わせでパスが存在する状態を理想とする。また、閾値として、スレッド名を用いる場合は重要度の上位10%をキーワードノードとして採用，スレッド本文を用いる場合は出現回数が5回未満1000回以上の単語を排除，重要度は得られる単語数の差を考慮し，上位0.1%とする。

表1 実験に用いるスレッド

スレッド名に含まれる語	スレッド数
youtube	33
ニコニコ動画	38
ニコニコ生放送	27

4.2 実験結果・考察

前節の条件で2部グラフの構築を行い，それぞれのグループにおけるパスの構築率を算出した結果，表2,3及び図2,3のようになった。図は枠に囲まれたものがキーワードノード，番号を囲む円がスレッドノードを表し，番号による色分けとグループが対応している。結果に対する考察は以下になる。

- ・スレッド名の場合では単純に出現頻度の高いキーワードが重要度の上位となる関係で，各グループに含まれる語を中心とするクラスが形成されるのが確認できた。一方で今回の結果では英文に見られる大文字小文字等の表記揺れが含まれており，その判別，修正が必要であると判明した。
- ・スレッド本文の場合では”ニコニコ生放送”のグループにおいて，スレッド名のときよりも高いパス構築率が確認された。この理由としては，グループ内で頻出かつ独自性・専門性の高い単語が多く，それによりとりわけ高い重要度を持つ単語を複数抽出できたためと考えられる。

表2 実験結果(スレッド名，キーワード数19)

	youtube	ニコニコ動画	ニコニコ生放送
孤立スレッド数	2	7	9
リンク数	88	73	51
パス構築率	88.1%	66.0%	51.0%

表3 実験結果(スレッド本文，キーワード数42)

	youtube	ニコニコ動画	ニコニコ生放送
孤立スレッド数	10	14	5
リンク数	101	89	99
パス構築率	47.9%	39.3%	77.0%

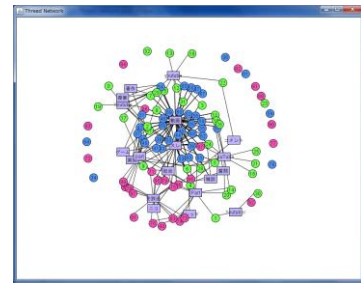


図2 2部グラフ構築結果(スレッド名)

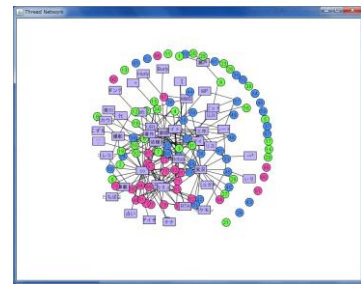


図3 2部グラフ構築結果(スレッド本文)

5. おわりに

本研究ではスレッドフロート型掲示板のスレッドを，スレッド名及び本文から重要語を抽出，スレッドとキーワードによる2部グラフの構築手法を提案した。

実験では理想解を設定した上で，スレッドに対し提案手法を適用し前述の結果が得られたが，実際のスレッドは実験で使用したものよりも遥かに多く，名称・内容共にその傾向も千差万別である。また，2ちゃんねるのような掲示板は独自の表現やスラング，表記揺れ等が多く，今回使用した辞書ではそれをカバー出来ていない。したがって実験結果について，その精度は保証が難しい。

今後の課題としては，形態素解析周りの前述する問題点への対処や2部グラフの表示の改善，及び構築した2部グラフに対するコミュニティ抽出によるスレッドの解析などが挙げられる。

参考文献

- [1]深谷 雅志, 倉本 到, 渋谷 雄, 辻野 嘉宏, “電子掲示板における行動履歴を用いたユーザにとって興味あるスレッドの推薦手法”, 電子情報通信学会技術研究報告, Vol.106, No.410, pp.149-154(2006).
- [2]MeCab (和布蕪), <http://mecab.sourceforge.net/>
- [3]Salton, G. and McGill, M.: *Introduction to Modern Information Retrieval*., McGraw-Hill Book Company (1984).
- [4]2ちゃんねる, <http://www.2ch.net/>