























One promising future direction for our method would be to assign a confidence score for each Y2H interaction (i.e. conservation of the interaction across species). It is possible to define a distance-based measure between proteins and use a Diffusion Map for spectral clustering, as in <sup>18</sup>. However, this method is very computationally expensive and hard to scale to large datasets. We plan to explore an efficient implementation of a continuous approach of diffusion maps with discrete approach of Gomory-Hu trees.

## 8. Conclusions

We have presented an efficient algorithm for identifying protein complexes through manipulation of the Gomory-Hu tree of the PPI Y2H network. Our method is shown to be robust against high FP and FN rates and capable of producing clusters of high quality when compared to other approaches. Identified functional modules are strong candidates for complex predictions and constitute reliable material for experimental research.

## References

1. The MIPD database: <http://mips.gsf.de/>
2. The BioGRID database: <http://www.thebiogrid.org/>
3. R. Sharan, T. Ideker, B. Kelley, R. Shamir, B. Karp, *Recomb.* 282–289 (2004).
4. A. Gavin, M. Bosche and R. Krause, *Nature.* **415**, 141–147 (2002).
5. Y. Ho, A. Gruhler and A. Heilbut, *Nature.* **415**, 180-183 (2002).
6. R. Gomory and T. Hu, *J. of SIAM.* **4**, 551-570 (1961).
7. R. Jansen, H. Yu, et al *Science.* **302(5644)**, 449-53 (2003).
8. I. Lee, S. Date, A. Adai, E. Marcotte, *Science.* **306(5701)**, 1555-1558 (2004).
9. G. Bader and C. Hogue, *BMC Bioinformatics.* bf 4(1), 2 (2003).
10. D. Matula, *FOCS.* 249–251 (1987).
11. H. Nagamochi and T. Ibaraki, *SIAM J.Disc. Math.* **5**, 54–66 (1992).
12. A. King, N. Przulj, I. Jurisica, *Bioinformatics.* **20(17)**, 3013–3020 (2004).
13. V. Spirin and L.A. Mirny, L.A., *Proc Natl Acad Sci.* **100**, 12123–12128 (2003).
14. S Maslov and K. Sneppen, *Science.* **296**, 910-913 (2002).
15. M. Newman, *Phys. Rev. E.* **74**, 036104 (2006).
16. E. Hartuv and R. Shamir, *IPL.* **76(4-6)**, 175–181 (2000).
17. G. Flake, R. Tarjan, K. Tsioutsoulouklis, *Internet Math.* **1(4)**, 385-408 (2004).
18. G. Lerman and B.E. Shakhnovich, *PNAS.* **104(27)**, 11334-11339 (2007).
19. J. Chen and B. Yuan, *Bioinformatics.* **22(18)**, 2283–2290 (2006).
20. E. Hirsh and R. Sharan, *Bioinformatics.* **23(2)**, e170-e176 (2007).
21. M. Meila. *COLT.* (2003).
22. S. Brohee and J. Van Helden, *BMC Bioinformatics.* **7:488** (2006).
23. J. Enright, S. Dongen, A. Ouzounis, *Nucl. Acids Res.* **30(7)**, 1575-84 (2002).
24. F. Luo, Y. Yang, et al *Bioinformatics.* **23(2)** 207-14 (2007).