**Статья**: 2012 - Heyne et al. - GraphClust alignment-free structural clustering of local RNA secondary structures

Для успешного понимания доклада необходимо узнать или вспомнить:

1. что такое non-coding RNA ([ncRNA](http://en.wikipedia.org/wiki/Non-coding_RNA)) и почему её нужно изучать;
2. проблему multiple sequence alignment ([MSA](http://en.wikipedia.org/wiki/Multiple_sequence_alignment)) и понять почему это [сложно](http://en.wikipedia.org/wiki/Multiple_sequence_alignment#Dynamic_programming_and_computational_complexity) и плохо;
3. что означает, что два графа [изоморфны](http://en.wikipedia.org/wiki/Graph_isomorphism), и почему эта задача NP-[полна](http://en.wikipedia.org/wiki/Graph_isomorphism_problem);

Опционально можно ознакомиться

1. с тем, как [устроены](http://www.cs.ucsb.edu/~xyan/tutorial/KDD08_graph_partII.pdf) graph kernels и зачем они нужны;
2. c алгоритмом MinHash, например, на [Википедии](http://en.wikipedia.org/wiki/MinHash) или в [статье](http://www.tonicebrian.com/2013/03/11/introduction-to-the-minhash-algorithm/) Toni Cebrián;
3. и с [различными](http://homepages.ulb.ac.be/~dgonze/TEACHING/stat_scores.pdf) метриками похожести биологических строчек: score, bit-score, P-value, E-value.

Обе темы будут изложены в ходе доклада, но полезно ознакомиться с ними заранее, чтобы помочь докладчику в случае затруднений.