

Задание 7. Алгоритмы на графах. Инструменты для анализа сетей

Варвара Кошман, С4113, 26.12.2019

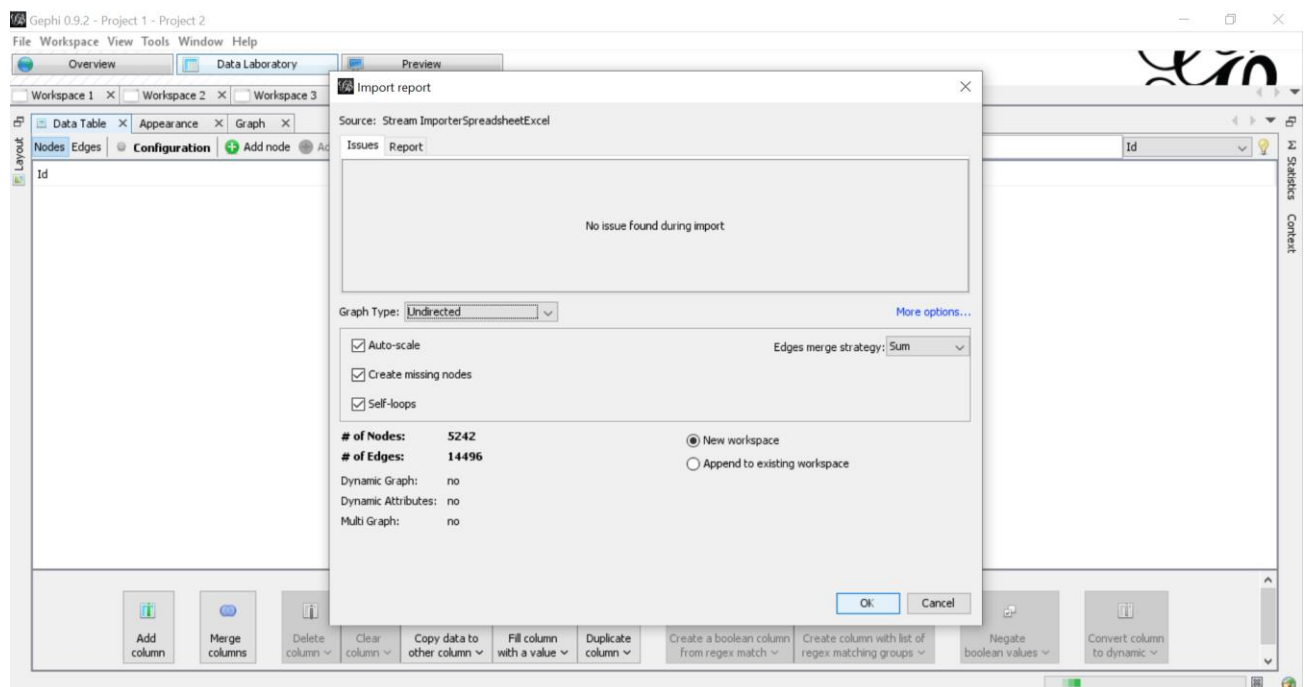
В качестве сети для анализа была выбрана сеть ученых-коллабораторов, которые публиковали статьи по теме общей теории относительности и квантовой космологии в arXiv. Если автор i написал в соавторстве статью с автором j , то граф содержит неориентированное ребро $i-j$. Если в статье k ученых указаны в соавторстве, то сеть генерирует полный подграф с k вершинами.

Данные были загружены с <http://snap.stanford.edu/data/ca-GrQc.html> в формате .txt, собственным программным кодом переформатированы и затем приведены в .csv.

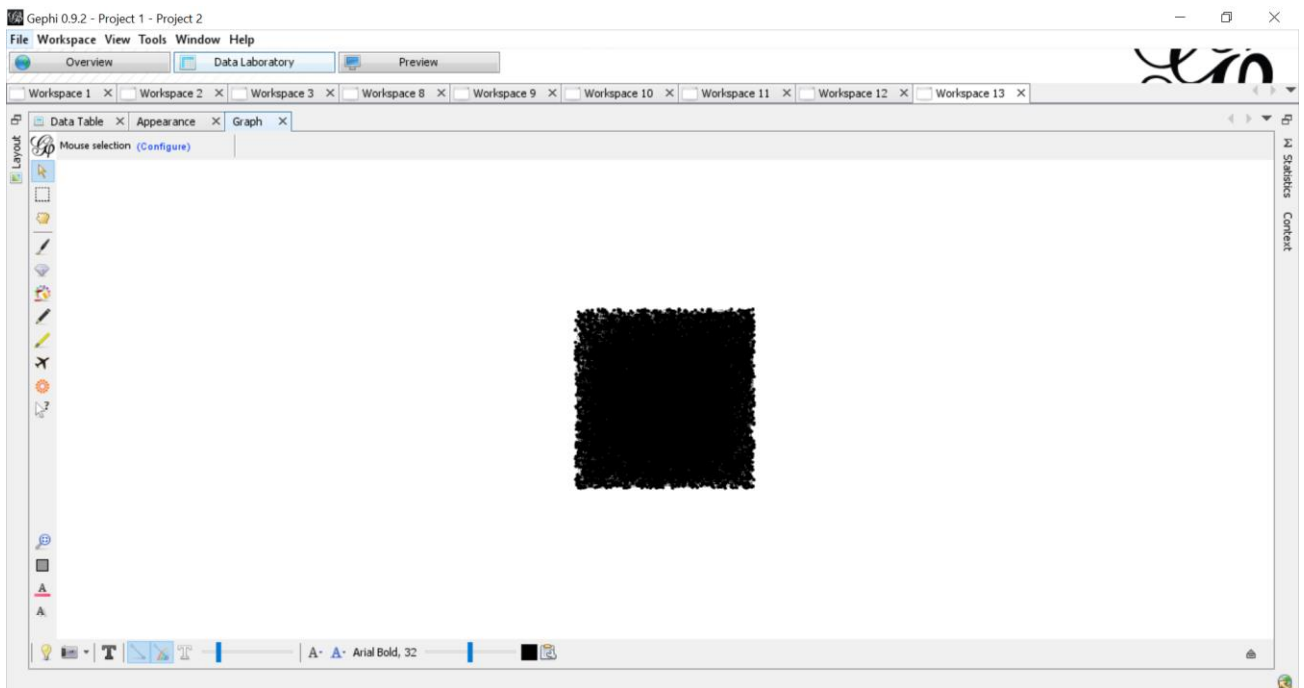
Данная сеть представляется в виде неориентированного, невзвешенного графа с 5242 вершинами и 14484 ребрами.

Некоторые основные шаги в процессе работы с Gephi:

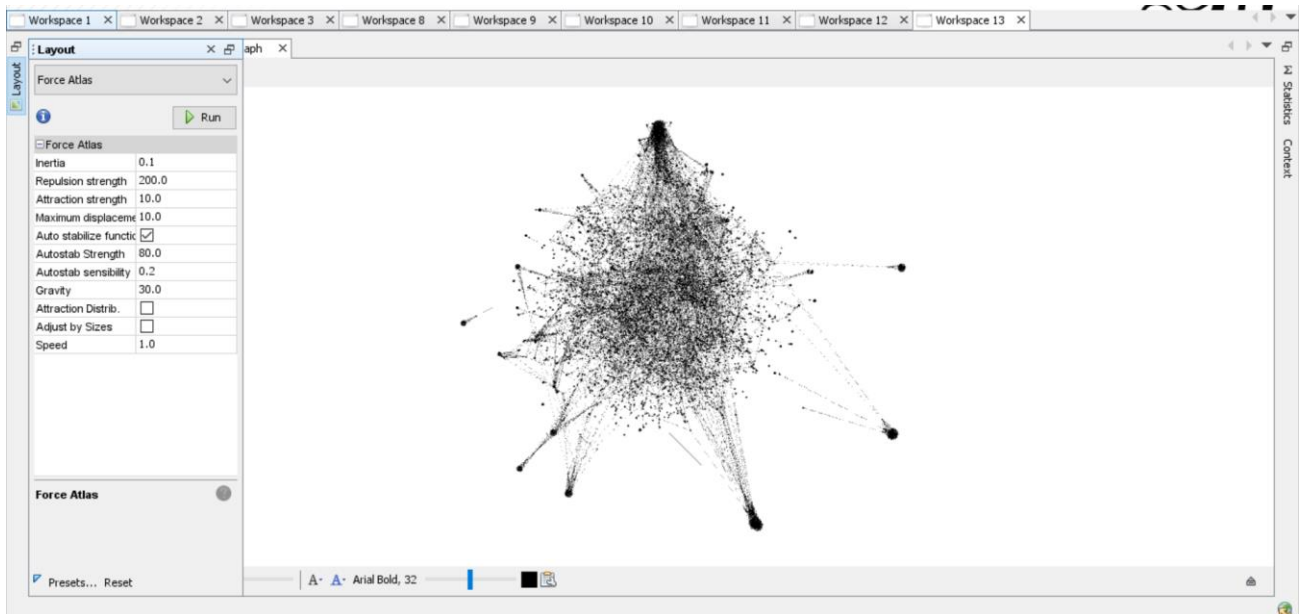
(импорт данных без ошибок)



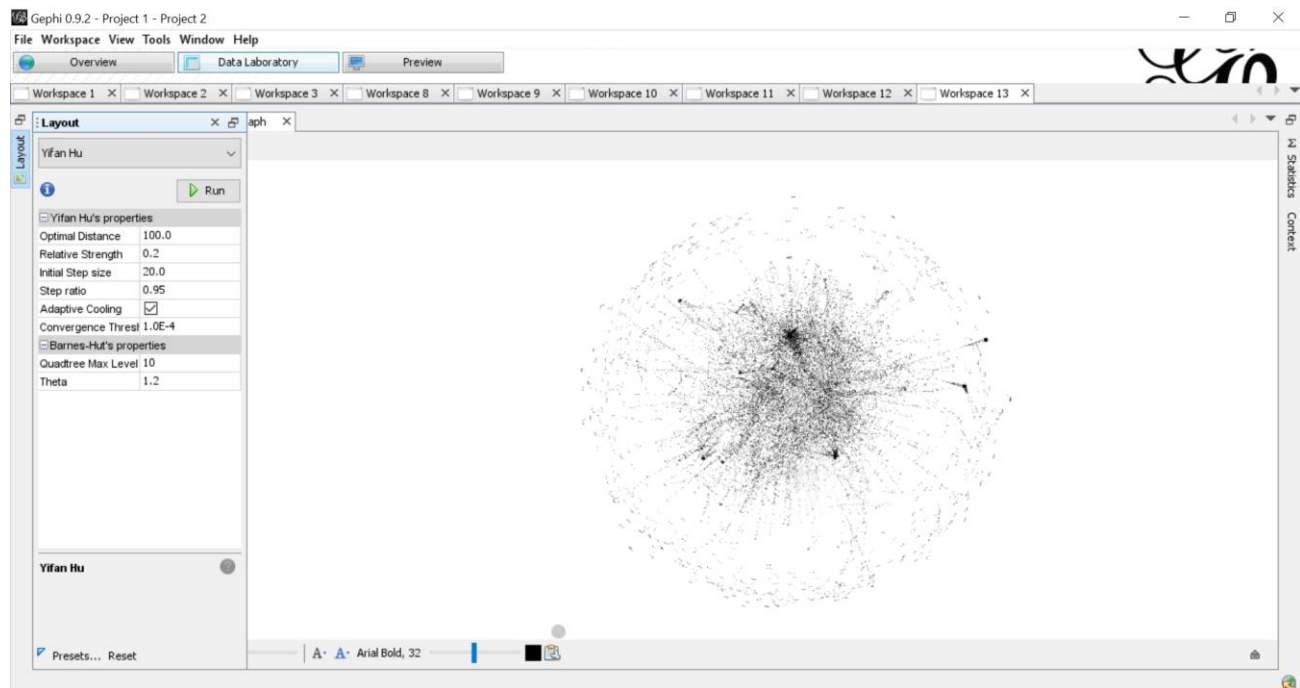
Первоначальное представление сети не информативно:



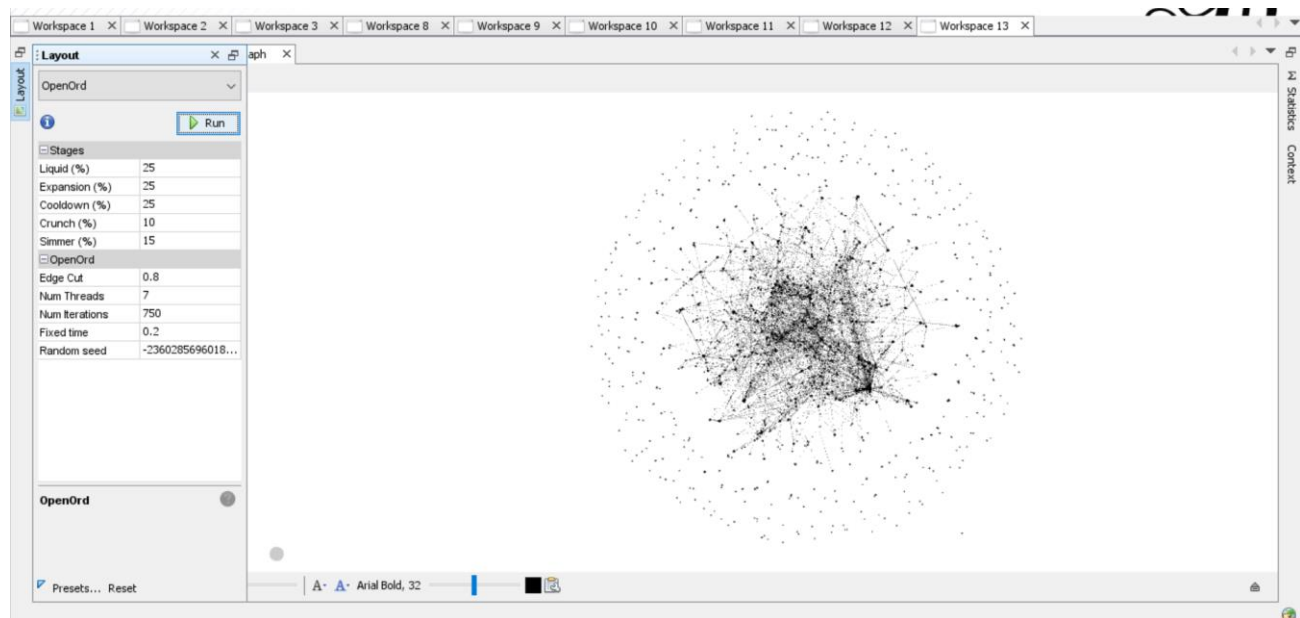
Меняем представление на **Force Atlas**:



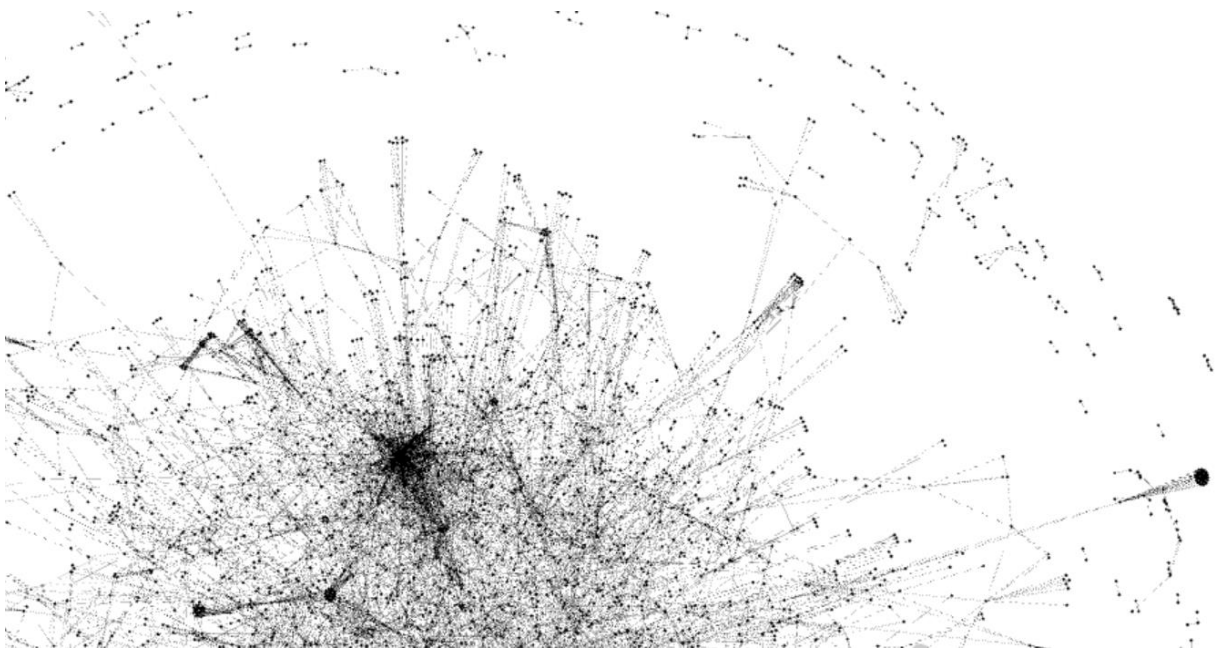
И на Yifan Hu:



На OpenOrd:



(вид ближе):



Из выбранных представлений видны выделяющиеся связанные кластеры, а так же изолированные: певрое означает, что некоторое число ученых в соавторстве писало много совместных статей и при этом многие из них писали статьи и с другими учеными, не из этого круга. Изолированные подграфы означают группу ученых, которые, кроме как друг с другом, статей на эту тему не публиковали.

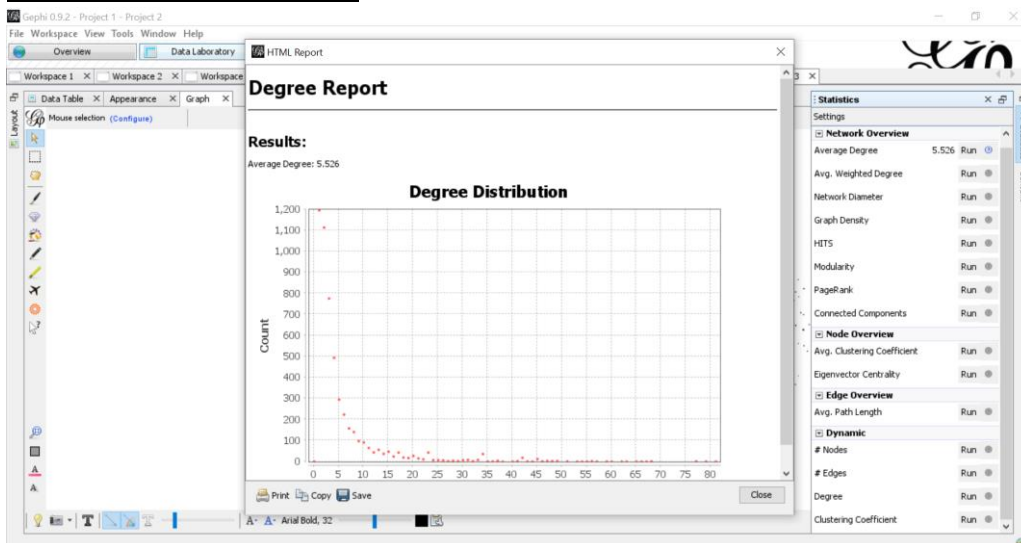
Подсчет статистики

Основные меры:

$$|V| = 5242$$

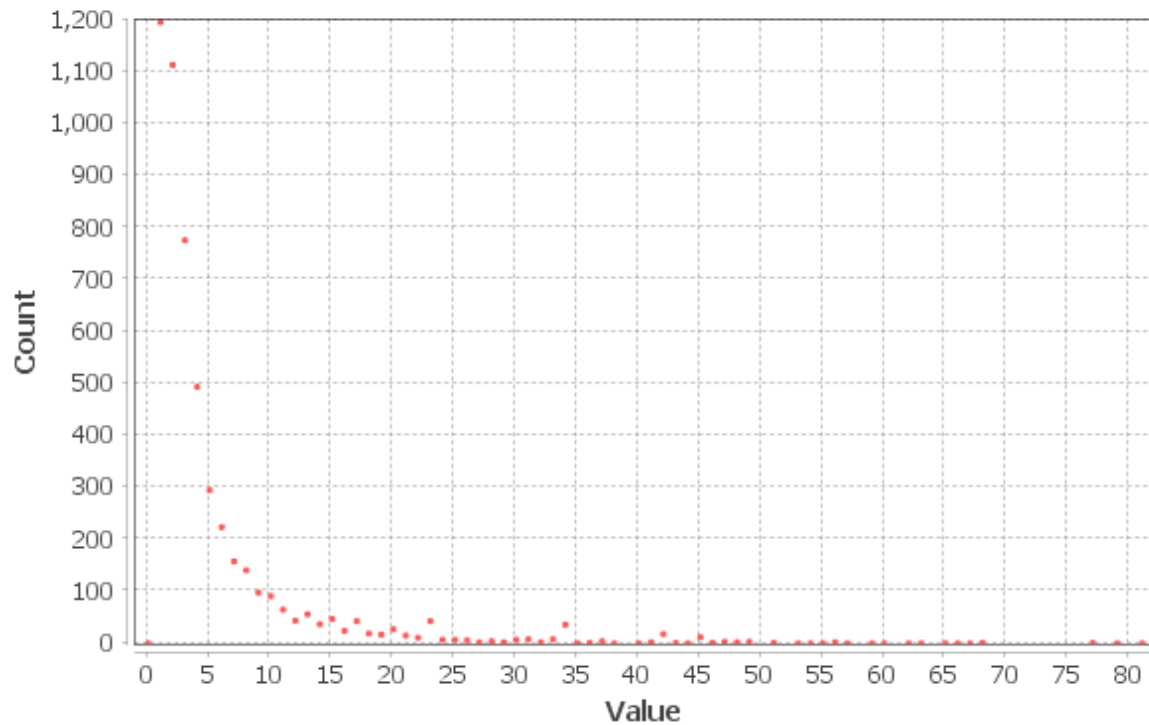
$$|E| = 14484$$

Распределение степеней:



Средняя степень вершин : 5,526

Degree Distribution



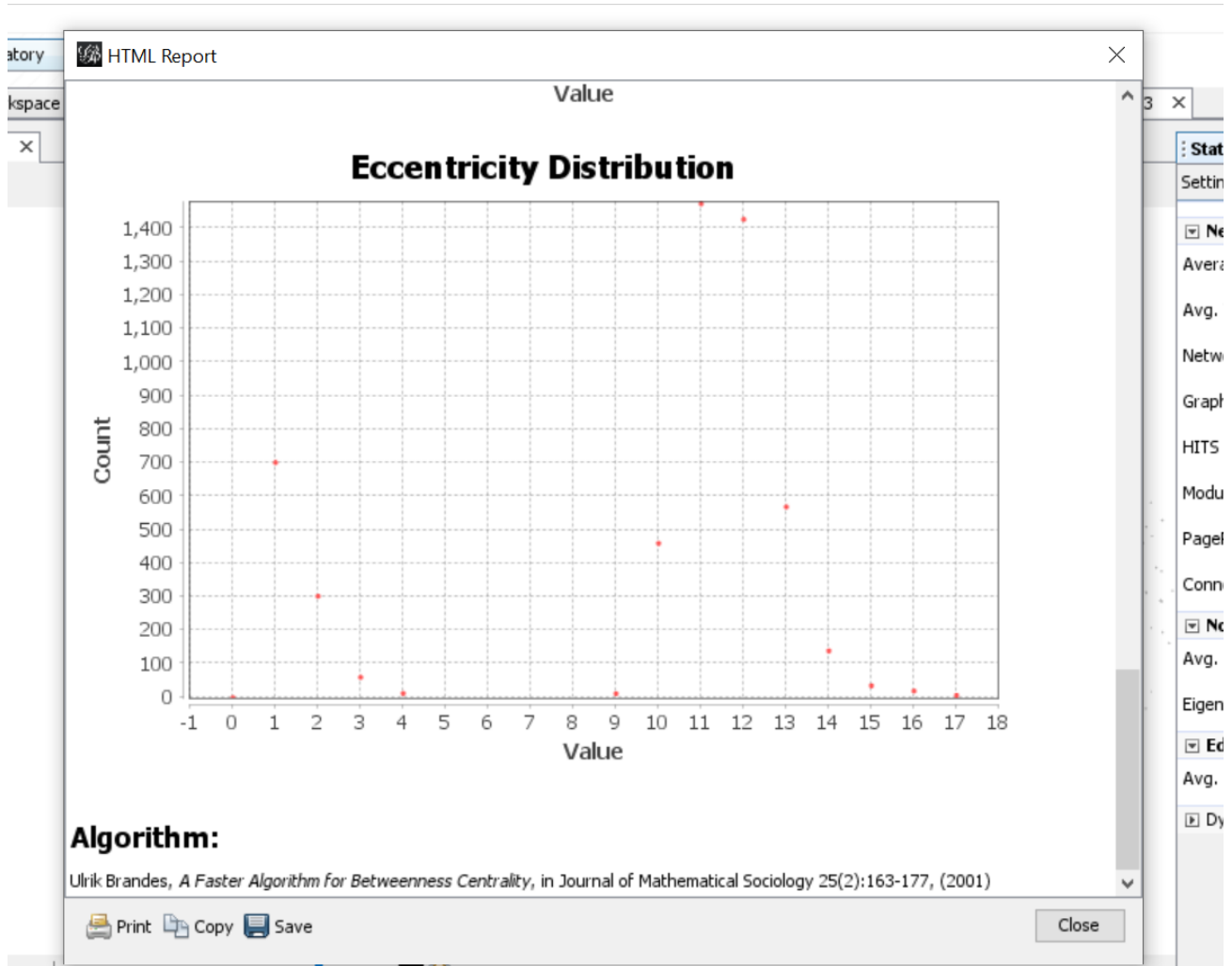
По графику видно, что большая часть ученых имеет от 1 до 10 связей с другими учеными по публикациям. И в среднем каждый имеет 5,526 связей с другими.

Меры расстояния:

Диаметр $D = 17$ (максимальное расстояние между парой вершин)

Радиус $r = 0$ (минимальный эксцентриситет по всем вершинам) (ученые, которые писали статьи сами и в соавторстве не работали)

Средняя длина пути $l = 6.048514960757815$



Мера плотности:

$\rho = 0,001$ – граф разреженный

Модулярность:

Modularity: 0.856 – граф содержит плотные внутрикластерные связи и разреженные межкластерные связи

Modularity with resolution: 0.856

Number of Communities: 387

Это значит, что ученые много публиковались по теме квантовой космологии устоявшимися группами и меньше с другими учеными, причем, этих устоявшихся сообществ довольно много.

Results:

Modularity: 0.856

Modularity with resolution: 0.856

Number of Communities: 387

Size Distribution