



ДИНАМИЧЕСКАЯ СЕТЬ ПОНЯТИЙ ИЗ ВЕБ-ПУБЛИКАЦИЙ

д.т.н. Ландэ Д.В., д.ф.-м.н. Снарский А.А.

Информационный центр «ЭЛВИСТИ»,
НТУУ «КПИ»

Одесса-2008



ПРОБЛЕМЫ ПРИКЛАДНОЙ ЛИНГВИСТИКИ ДИНАМИЧЕСКАЯ СЕТЬ ПОНЯТИЙ ИЗ ВЕБ-ПУБЛИКАЦИЙ

Характеристики информационного потока:

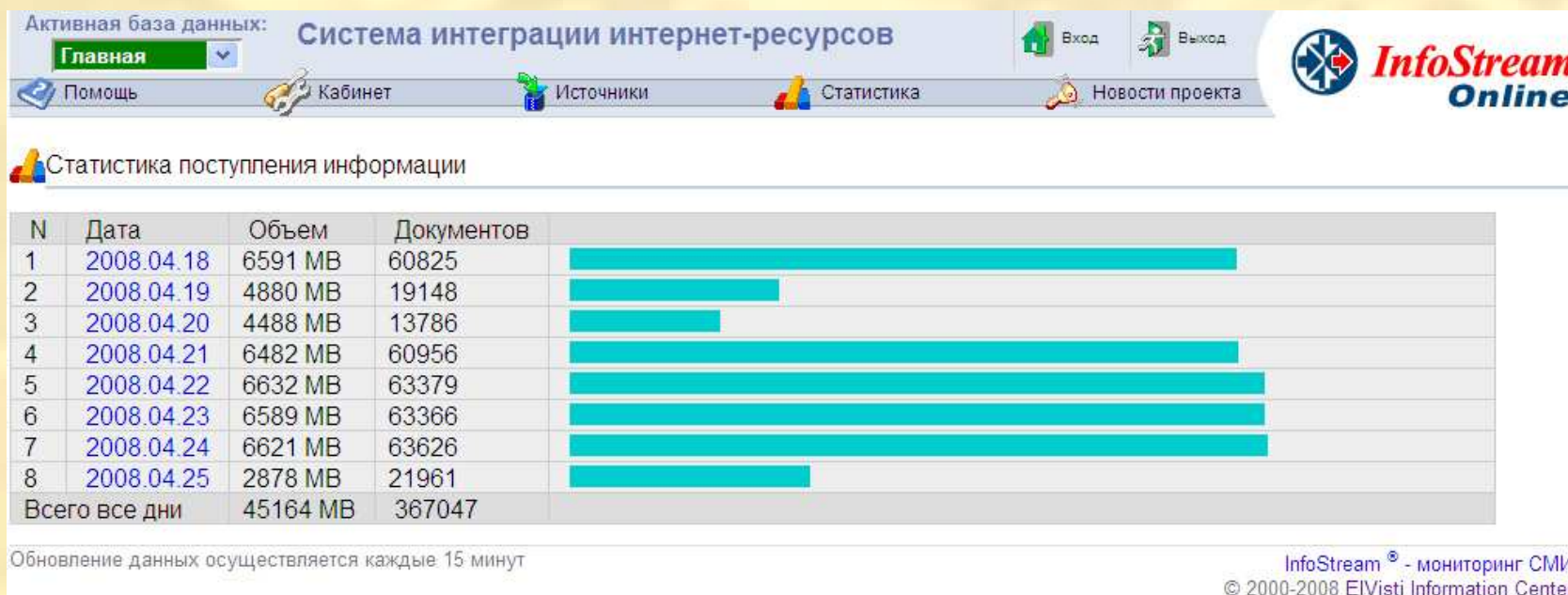


Более 3 300 источников информации

Более 1 200 000 сообщений в месяц

Более 60 000 000 документов в ретрофонде

Обновление базы данных – каждые 15 минут





ПРОБЛЕМЫ ПРИКЛАДНОЙ ЛИНГВИСТИКИ ДИНАМИЧЕСКАЯ СЕТЬ ПОНЯТИЙ ИЗ ВЕБ-ПУБЛИКАЦИЙ

Экстрагированные
понятия
(персоны) в
информационном
портрете:

Активная база данных: Система интеграции интернет-ресурсов

Вход Выход

InfoStream Online

Помощь Кабинет Источники Статистика Новости проекта

одесс припорт

Период: 7 дней Убрать дубли Морфология

Найти Динамика Дайджест

Очистить События Сюжеты

Язык запросов Примеры

одесс припорт

Найдено документов - 658, страница 1 из 44

Статистика слов:

ОДЕСС - 10475, ПРИПОРТ - 2112

Добавить канал

Информационный портрет

Уточнить запрос

Рубрики (26)

Языки (2)

Страны источников (7)

Источники (50)

Размер (3)

Цифровая насыщенность (3)

Тональность (2)

География (50)

Персоны (50)

AND	NOT
<input type="checkbox"/> Ющенко ***	<input type="checkbox"/>
<input type="checkbox"/> Тимошенко **	<input type="checkbox"/>
<input type="checkbox"/> Семенов-самсоненко *	<input type="checkbox"/>
<input type="checkbox"/> Турчинов *	<input type="checkbox"/>
<input type="checkbox"/> Балага	<input type="checkbox"/>
<input type="checkbox"/> Янукович	<input type="checkbox"/>
<input type="checkbox"/> Самсоненко	<input type="checkbox"/>
<input type="checkbox"/> Портнов	<input type="checkbox"/>
<input type="checkbox"/> Писаренко	<input type="checkbox"/>
<input type="checkbox"/> Кириленко	<input type="checkbox"/>
<input type="checkbox"/> Томенко	<input type="checkbox"/>
<input type="checkbox"/> Дубин	<input type="checkbox"/>
<input type="checkbox"/> Ехануров	<input type="checkbox"/>
<input type="checkbox"/> Пукшин	<input type="checkbox"/>
<input type="checkbox"/> Яценюк	<input type="checkbox"/>
<input type="checkbox"/> Райков	<input type="checkbox"/>
<input type="checkbox"/> Бойко	<input type="checkbox"/>
<input type="checkbox"/> Луценко	<input type="checkbox"/>
<input type="checkbox"/> Плачков	<input type="checkbox"/>
<input type="checkbox"/> Фирташу	<input type="checkbox"/>

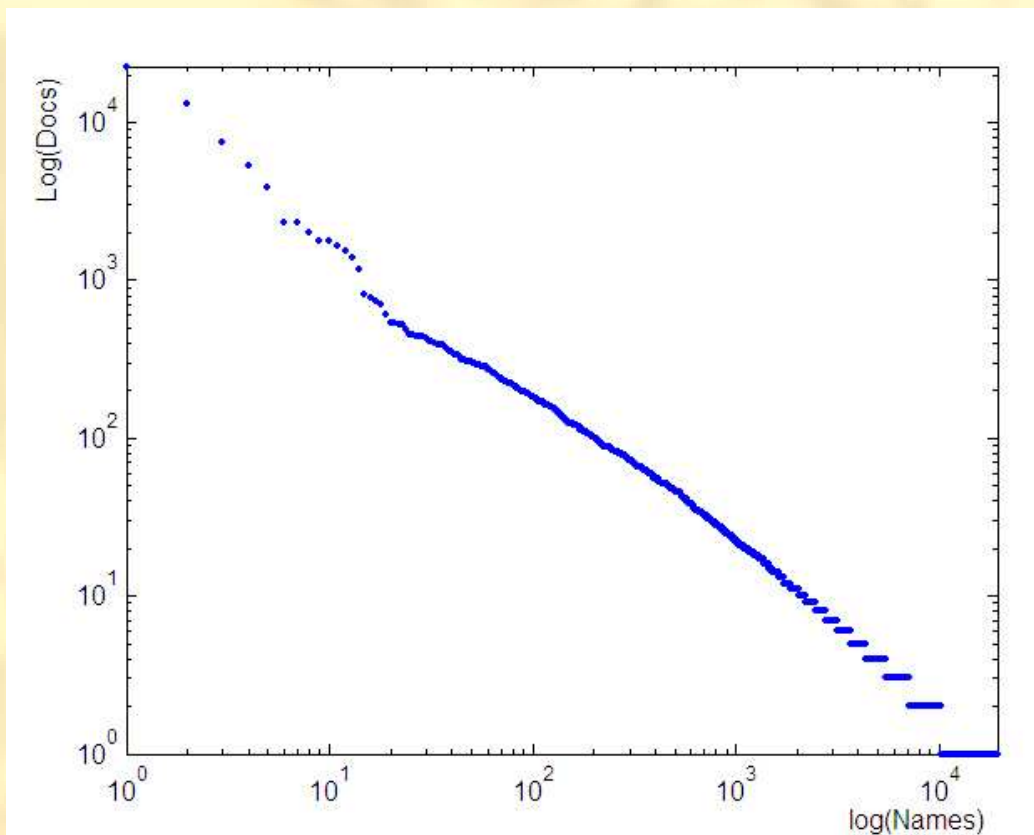
- Суд приостановил указы Ющенко о кадровых назначениях в ФГИ**
ИнтерМедиа Консалтинг 2008.04.25 12:30
Суд приостановил указы Президента о кадровых назначениях в Фонде госимущества и приостановке приватизации Одесского припортового завода - Портнов. УНИАН
Похожие документы - Оригинал
- Поменять Семенов Тимошенко позволил суд**
4POST 2008.04.25 12:30
24 апреля окружной административный суд города Киева приостановил действие Указа Президента о восстановлении Валентины Семенов-Самсоненко в должности главы Фонда госимущества. Об этом сообщила премьер-министр Украины Юлия Тимошенко журналистам на пресс-конференции.
Похожие документы - Оригинал
- Суд приостановил указы Ющенко о кадровых назначениях в ФГИ**
"2000" Новости 2008.04.25 12:30
Суд приостановил указы Президента о кадровых назначениях в Фонде госимущества и приостановке приватизации Одесского припортового завода - Портнов.
Похожие документы - Оригинал
- Суд приостановил указы президента относительно решений правительства о смене руководства ФГИ, а также приказ Фонда о приостановке приватизации ОПЗ - Тимошенко**
Интерфакс-Украина 2008.04.25 12:29
Премьер-министр Украины Юлия Тимошенко заявляет, что суд приостановил указы президента относительно руководства Фонда государственного имущества (ФГИ), а также все приказы главы ФГИ о приостановке приватизации Одесского припортового завода.
Похожие документы - Оригинал
- Суд приостановил указы Президента по ФГИ и все приказы главы ФГИ о приостановке приватизации ОПЗ - Тимошенко**
ISTV 2008.04.25 12:27
Премьер-министр Украины Юлия Тимошенко заявляет, что суд приостановил указы Президента относительно руководства Фонда государственного имущества, а также все приказы главы ФГИ о приостановке приватизации Одесского припортового завода.
Похожие документы - Оригинал
- По сценарию катастрофы**
Свобода 2008.04.25 12:18
Есть есть пророки в своем Отечестве! Сошлюсь на статью руководителя Фонда социально-политических инициатив Романа Василюшина и народного депутата Украины IV созыва Александра Ткаленко, опубликованную под выразительным заголовком "Чому Ющенкові не треба обиратися вдруге" еще в прошлом году, в момент, когда Президент вынужден был внести кандидатуру Тимошенко на должность премьер-министра.
Похожие документы - Оригинал
- Валентину Семенов обвинили в государственном рейдерстве**



ПРОБЛЕМЫ ПРИКЛАДНОЙ ЛИНГВИСТИКИ ДИНАМИЧЕСКАЯ СЕТЬ ПОНЯТИЙ ИЗ ВЕБ-ПУБЛИКАЦИЙ

Описываются свойства сетей, образуемых понятиями, которые связаны друг с другом упоминанием в одних и тех же документах. Исследовалась сеть, образуемая фамилиями персон, извлекаемых из текстовых корпусов интернет-СМИ по общеполитической тематике 1 месяц объемом 55 тысяч документов. Всего в текстах упоминалось свыше 19 тыс. персон.

22637	name . Ющенко
12986	name . Янукович
7468	name . Тимошенко
5305	name . Мороз
3890	name . Путин
2311	name . Кириленко
2298	name . Симоненко
2001	name . Луценко
1768	name . Балого
1756	name . Пшеничный
1635	name . Цушко
1526	name . Азаров
1383	name . Турчинов
1180	name . Кучма
803	name . Гайдук
778	name . Гавриш
699	name . Иванов
599	name . Кравчук
531	name . Миронов
530	name . Яценюк
522	name . Гриценко
483	name . Черновол
452	name . Березовский
449	name . Киселев
443	name . Чечетов
441	name . Волга
...	

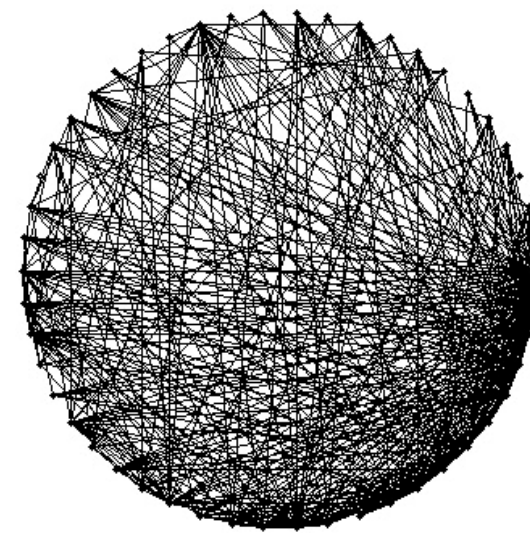
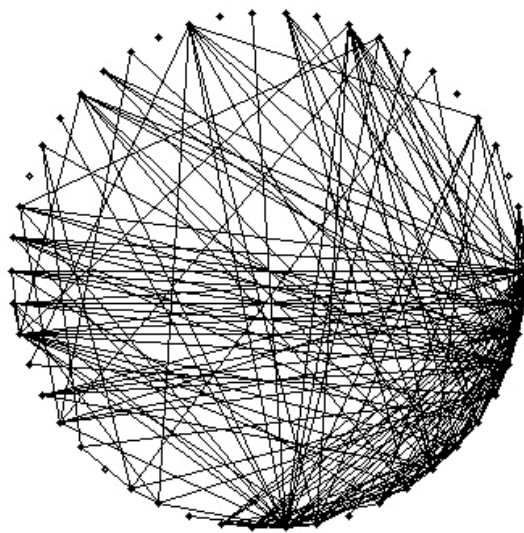
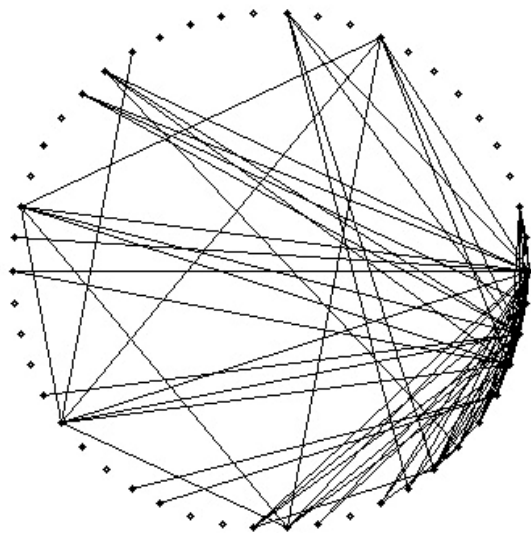




ПРОБЛЕМЫ ПРИКЛАДНОЙ ЛИНГВИСТИКИ ДИНАМИЧЕСКАЯ СЕТЬ ПОНЯТИЙ ИЗ ВЕБ-ПУБЛИКАЦИЙ

Исследуется сеть понятий, узлами которой выступают персоны, а ребра, соединяющие узлы, соответствовали количеству упоминаний соответствующих персон в одних и тех же документах.

Сеть, образуемая понятиями, извлекаемыми из потоков текстов, не является статической, а зависит от постоянно появляющихся новых документов, из которых извлекаются соответствующие понятия. Следовательно, для понимания структуры такой сети необходимо учитывать ее эволюцию.



Динамика развития сети

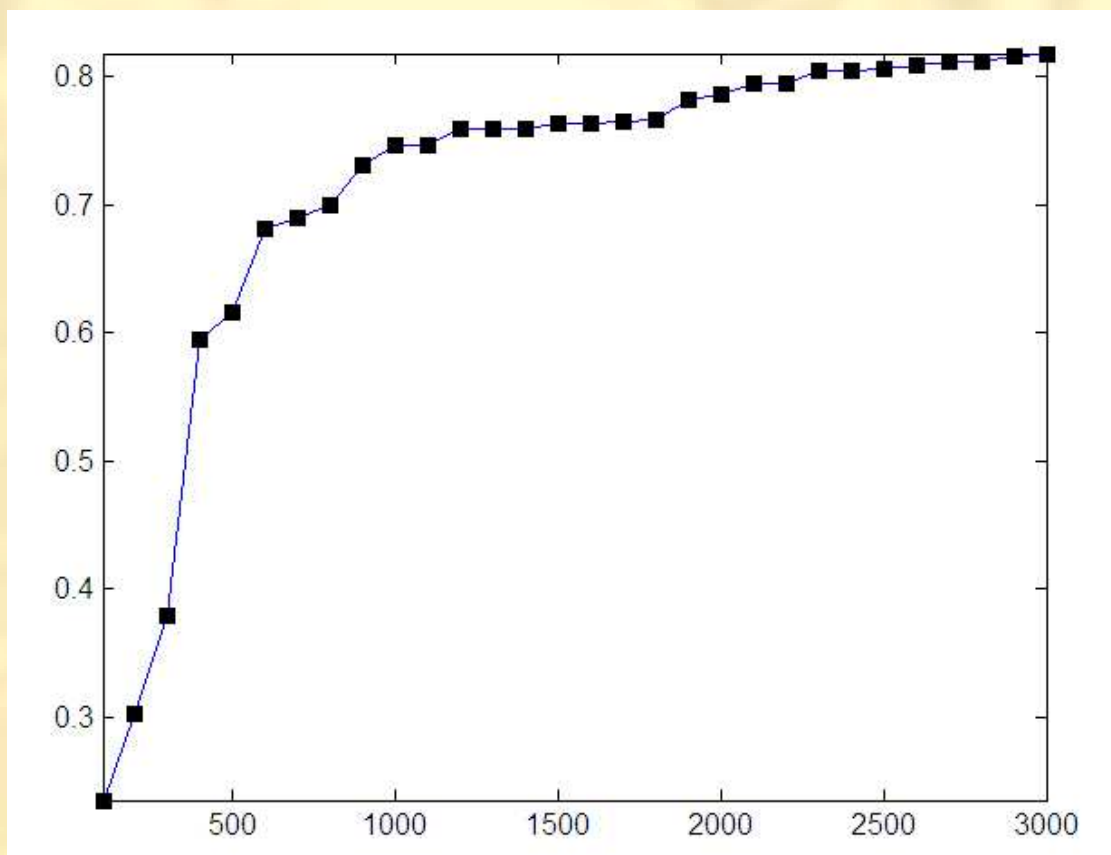


ПРОБЛЕМЫ ПРИКЛАДНОЙ ЛИНГВИСТИКИ ДИНАМИЧЕСКАЯ СЕТЬ ПОНЯТИЙ ИЗ ВЕБ-ПУБЛИКАЦИЙ

Средний инверсный путь между узлами:

$$il = \frac{2}{n(n-1)} \sum_{i>j} \frac{1}{d_{ij}},$$

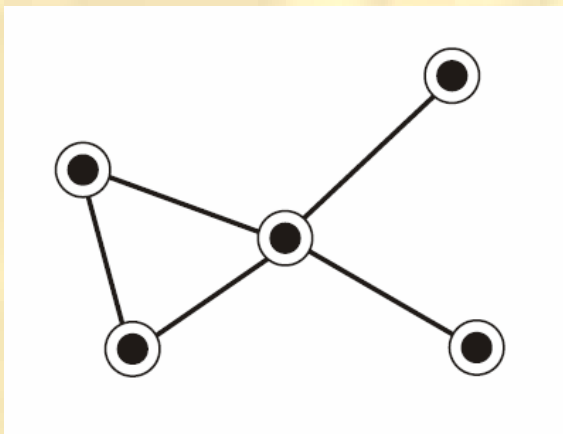
где d_{ij} – кратчайшее расстояние между узлами i и j .





ПРОБЛЕМЫ ПРИКЛАДНОЙ ЛИНГВИСТИКИ ДИНАМИЧЕСКАЯ СЕТЬ ПОНЯТИЙ ИЗ ВЕБ-ПУБЛИКАЦИЙ

Коэффициент кластерности для узлов, ранжированных по степени



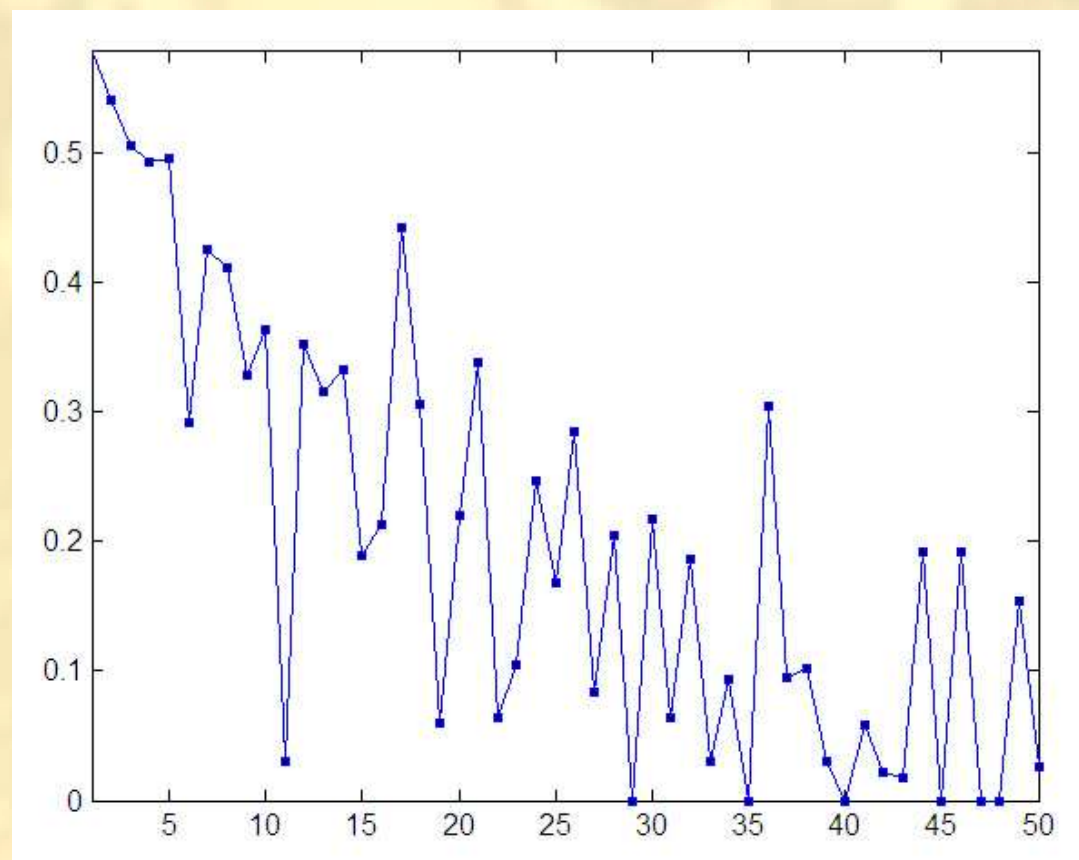
$c(0)$	= 0.422	- name.Ющенко
$c(1)$	= 0.459	- name.Янукович
$c(2)$	= 0.494	- name.Тимошенко
$c(3)$	= 0.507	- name.Мороз
$c(4)$	= 0.502	- name.Путин
$c(5)$	= 0.706	- name.Кириленко
$c(6)$	= 0.574	- name.Симоненко
$c(7)$	= 0.587	- name.Луценко
$c(8)$	= 0.670	- name.Балога
$c(9)$	= 0.636	- name.Пшеничний
$c(10)$	= 0.962	- name.Цушко
$c(11)$	= 0.647	- name.Азаров
$c(12)$	= 0.683	- name.Турчинов
$c(13)$	= 0.667	- name.Кучма
$c(14)$	= 0.808	- name.Гайдук
$c(15)$	= 0.785	- name.Гавриш
$c(16)$	= 0.688	- name.Иванов
$c(17)$	= 0.932	- name.Кравчук
$c(18)$	= 0.769	- name.Миронов
$c(19)$	= 0.660	- name.Яценюк
$c(20)$	= 0.929	- name.Гриценко
$c(21)$	= 0.752	- name.Чорновил
$c(22)$	= 0.827	- name.Березовский
$c(23)$	= 0.712	- name.Киселев
$c(24)$	= 0.913	- name.Чечетов



ПРОБЛЕМЫ ПРИКЛАДНОЙ ЛИНГВИСТИКИ

ДИНАМИЧЕСКАЯ СЕТЬ ПОНЯТИЙ ИЗ ВЕБ-ПУБЛИКАЦИЙ

Коэффициент «промежуточности» для узлов, ранжированных по степени:

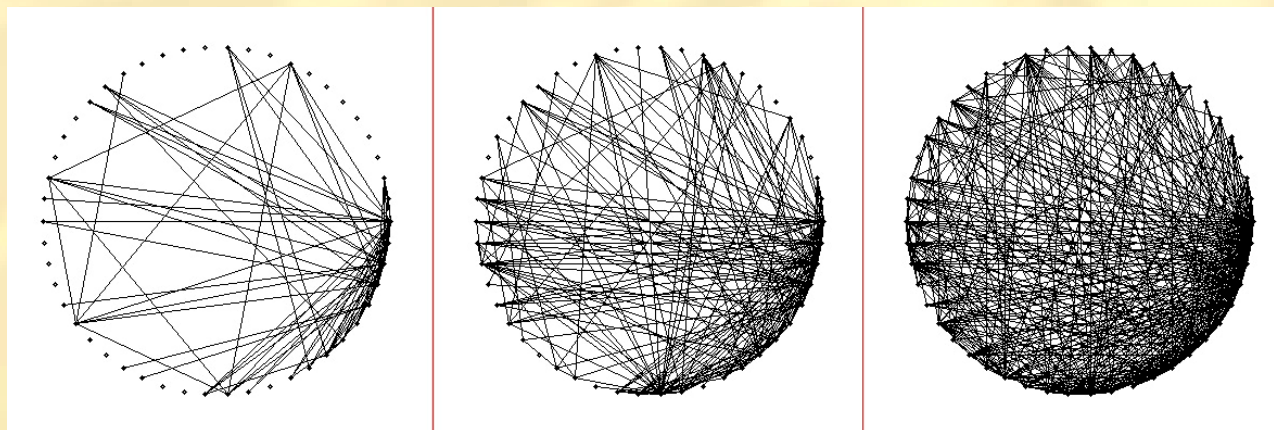
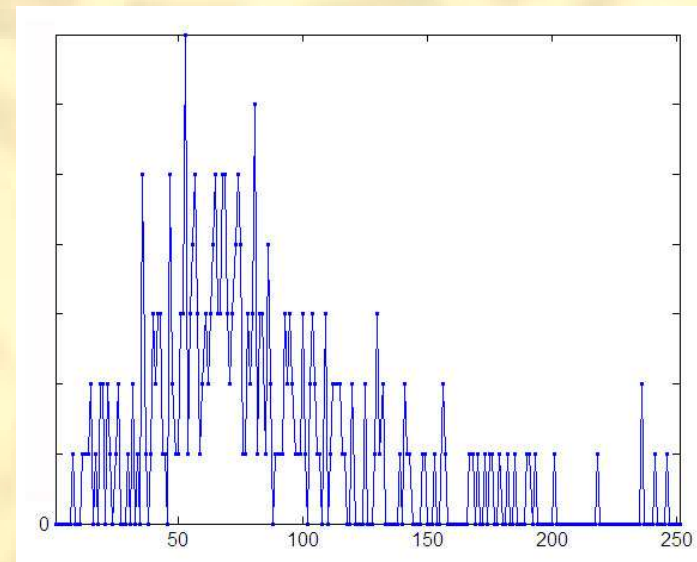
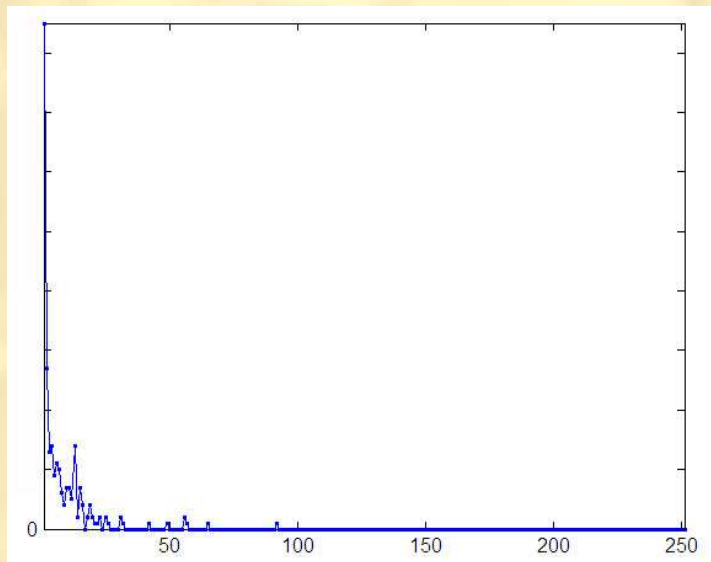




ПРОБЛЕМЫ ПРИКЛАДНОЙ ЛИНГВИСТИКИ

ДИНАМИЧЕСКАЯ СЕТЬ ПОНЯТИЙ ИЗ ВЕБ-ПУБЛИКАЦИЙ

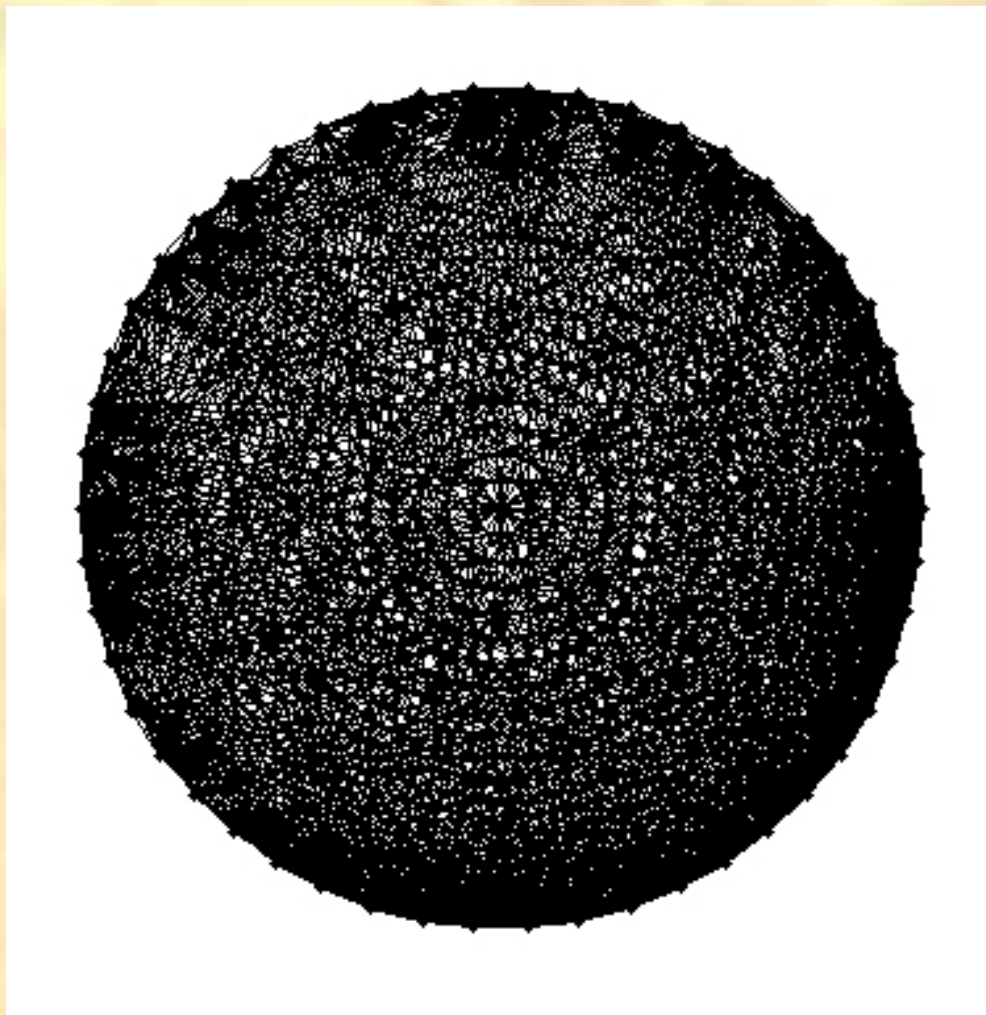
Распределение степеней узлов:





ПРОБЛЕМЫ ПРИКЛАДНОЙ ЛИНГВИСТИКИ ДИНАМИЧЕСКАЯ СЕТЬ ПОНЯТИЙ ИЗ ВЕБ-ПУБЛИКАЦИЙ

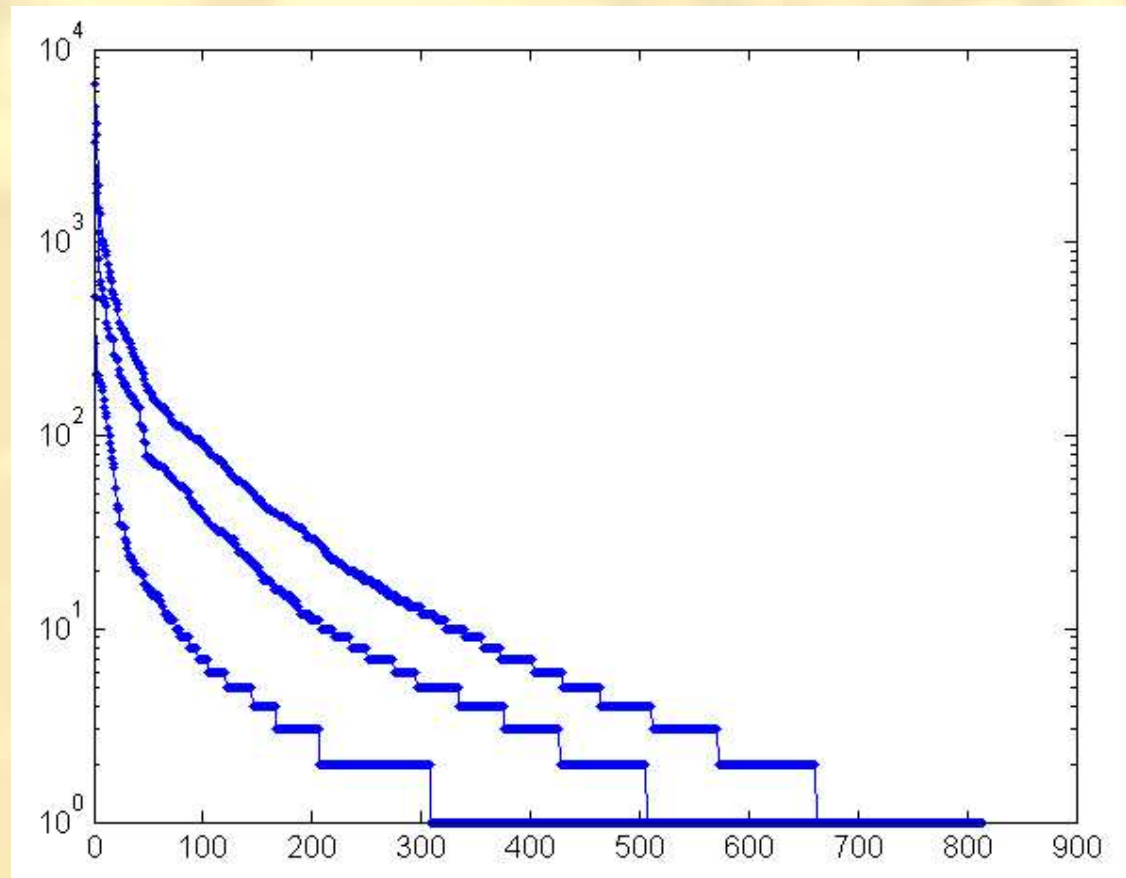
Сеть, близкая к состоянию деградации: 50 персон, 50000 документов:





ПРОБЛЕМЫ ПРИКЛАДНОЙ ЛИНГВИСТИКИ ДИНАМИЧЕСКАЯ СЕТЬ ПОНЯТИЙ ИЗ ВЕБ-ПУБЛИКАЦИЙ

Распределение весов ребер графа из 50 персон в логарифмической шкале (ось ординат) для массивов веб-публикаций из 1000, 10000 и 50000 документов (ось абсцисс – ранжированные по весам ребра):



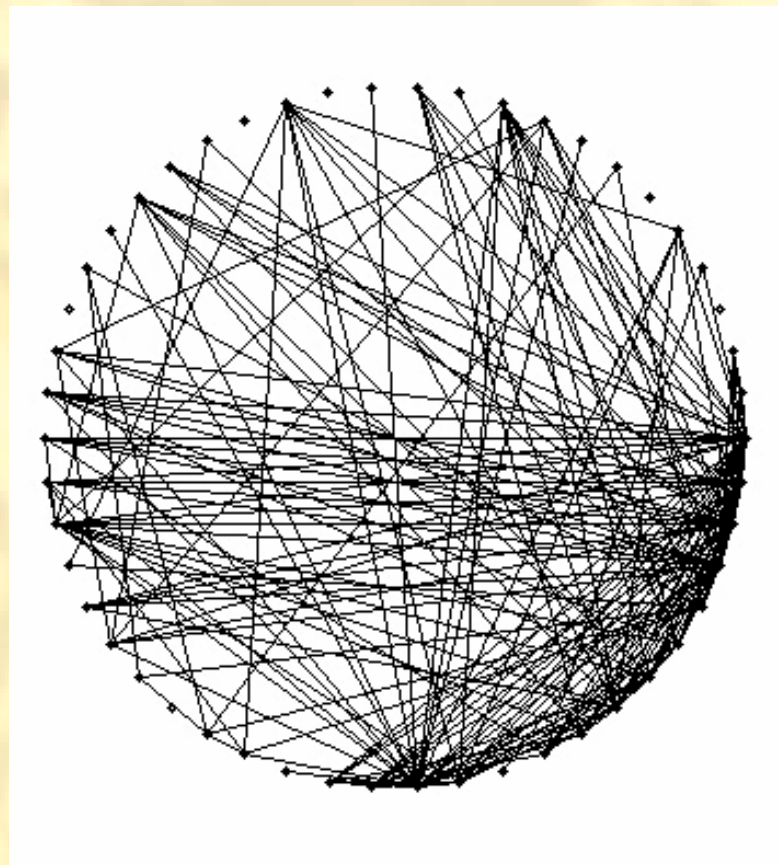


ПРОБЛЕМЫ ПРИКЛАДНОЙ ЛИНГВИСТИКИ ДИНАМИЧЕСКАЯ СЕТЬ ПОНЯТИЙ ИЗ ВЕБ-ПУБЛИКАЦИЙ

«Наложенная» сеть с весами ребер:

$$v' = \begin{cases} 1, & v \geq \varepsilon v_{\max} \\ 0, & v < \varepsilon v_{\max} \end{cases}$$

где v' - вес ребра наложенной сети,
 v - вес ребра исходной сети персон,
 v_{\max} - максимальное значение веса
ребер, ε - коэффициент закругления.



Феномен стабилизации наложенной сети:

для значения $\varepsilon=0.001$, 50 персон и количества документов в исходном массиве от 1000 до 50000, коэффициент кластерности составляет 0.78 ± 0.01 , а среднее инверсное расстояние - 0.65 ± 0.02 .



ПРОБЛЕМЫ ПРИКЛАДНОЙ ЛИНГВИСТИКИ

Спасибо за внимание!

Ландэ Дмитрий Владимирович

**Информационный центр «ЭЛВИСТИ»,
НТУУ «КПИ»**

stream@visti.net