Zero-Inflated Embeddings to Analyze Homicide Occurrence Patterns

Seminario de Matemáticas Aplicadas

Marzo, 2021

quantil

Results of the project "Diseño y validación de modelos de analítica predictiva de fenómenos de seguridad y convivencia para la toma de decisiones en Bogotá" funded by Colciencias with resources from the Sistema General de Regalías, BPIN 2016000100036. The opinions expressed are solely those of the authors.

Zero-Inflated Embeddings

- Analyzing homicide data is a challenging task due to its **low-frequency** and **spatial sparsity**.
- We use Zero Inflated Exponential Family Embeddings (ZIE) to analyze spatial patterns in Bogotá, Colombia.
- ZIE model provides useful insights about the different types of *cuadrantes* with an intuitive classification of high, medium, and low homicide-rate.

Related Literature

- Predictive models (Mohler et al.; Jin et al.) rely on large data sets to predict criminal events.
- Embedding methods have proven useful in identifying similar criminal dynamics in spatial and temporal units (Zhu et al.; Wang et al.; Yang et al.).

Data

- We focus our analysis on 1051 police jurisdictions called *cuadrantes*, made up in average by 10 - 15 blocks.
- We use the six weekly police patrol shifts used by the police department: morning, afternoon, and night shifts, differentiating between weekends and weekdays.
- Each observation is a count of homicides for each *cuadrante* in a police shift. Over the entire period for which data is available (2013-2019), this results in a very sparse matrix with less than 0.4% of non-zero entries.

Data

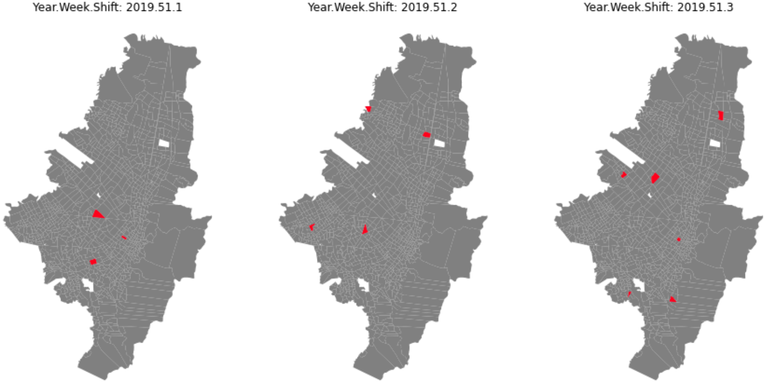


Figure: *Cuadrantes* in which homicides occurred during weeks shifts 1, 2, and 3 of a week in 2019.

Zero-Inflated Embeddings

- Exponential Family Embeddings (EFE) (Liu and Blei, 2017) aim to learn vector representations of items.
- For each item (spatial region) $j \in \{1, ..., J\}$, **learn embedding vector** $\rho_j \in \mathbb{R}^K$ **and context vector** $\alpha_j \in \mathbb{R}^K$ **from observation-context pairs**.
- Need three ingredients: context, conditional distribution and embedding structure.

Conditional Distribution

- Conditional distribution of $\boldsymbol{x}_i$ given its context $\boldsymbol{y}_i$

$$\boldsymbol{x} \sim \text{ExpFam}(\eta(\boldsymbol{y}_i, s_i), T(\boldsymbol{x}_i))$$

  with natural parameter through η and sufficient statistic T

$$\eta(\boldsymbol{y}_i, s_i) = f\left(\rho_{s_i}^{\top} \sum_{j \in c_i} y_{ij}\alpha_j\right).$$

- EFE conditional probability $p(\boldsymbol{x}_i | \boldsymbol{y}_i; \rho_{s_i}, \alpha_{c_i})$ captures the **interaction between items the observations in** s_i **and their context in** c_i: counts for all other *cuadrantes* in the city over the same period.

Exposure modeling

Zero-valued observations dominate: **embeddings focus on non-zeros** by modeling probability of being exposed to context.

- For each x_{ij} define $b_{ij} \sim \text{Bernoulli}(u_{ij})$
- Fit logistic regression: $u_{ij} = \text{logit}(\boldsymbol{w}_j^\top \boldsymbol{v}_i + w_j^0)$ with v_i vector of exposure covariates.
- $x_{ij} = \begin{cases} \delta_0 & b_{ij} = 0 \\ \text{ExpFam}(\eta(\boldsymbol{y}_i, s_i), T(\boldsymbol{x}_i)) & b_{ij} = 1 \end{cases}$

Implementation

- We fit low dimensional embeddings for each of the police *cuadrantes* where homicides have occurred for 2018 and 2019.
- We use Poisson (p) and Negative Binomial (nb) as exponential families, and we use $\log \text{softplus}(\cdot)$ as the link function f.
- The exposure covariates are: indicator for holiday, indicator for weekend, patrol shift, and month. We fit the embeddings using 80% of the data and test on the remaining 20%, corresponding to the most recent observations.
- The log-likelihood of observations given their contexts is maximized with gradient descent using AdaGrad. During training, 10% of the training data is used for validation of convergence.

Test Metrics

The Poisson zero-inflated embeddings using city context and temporal exposure covariates (ZIEC-p) has the best performance for different specifications of conditional distribution, context, and exposure covariates.

K		EFE-p	ZIE-p	ZIEC-p
8	all	-0.057(0.014)	**-0.040(0.006)**	-0.048(0.010)
	pos	-0.91(0.009)	-0.579(0.032)	**-0.459(0.100)**
16	all	-0.056(0.013)	**-0.040(0.006)**	-0.048(0.010)
	pos	-0.803(0.010)	-0.56(0.032)	**-0.441(0.083)**
32	all	-0.059(0.012)	**-0.043(0.006)**	-0.049(0.009)
	pos	-0.886(0.007)	-0.656(0.019)	**-0.484(0.066)**

Table: Predictive log-likelihood embeddings with Poisson exponential family.

Principal and Exposure Components

- Principal Component Analysis to investigate the resulting embeddings.
- First two PC explain 70.7% and 6.95% of the variance.
- Significant correlation between exposure probability and total homicide count.

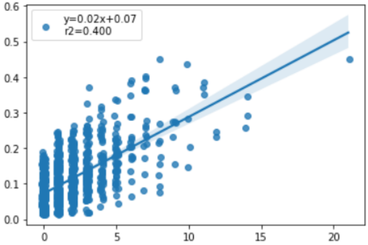
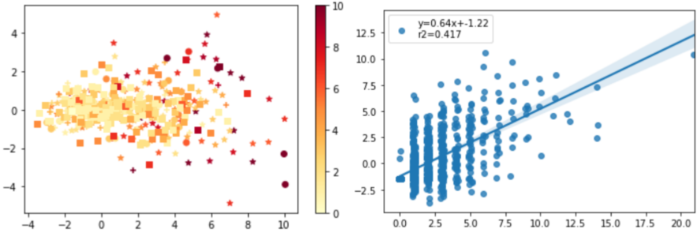


Figure: Correlation between exposure probability and total homicide count.

Principal and exposure components

Significant correlation between the first PC and total homicide count: ***Cuadrantes* where homicides occur more often cluster together.**



(a) PCA plot of the embedding (b) Correlation between first PC and total homicide count

Figure: Analyzing the role of the total homicide count on the embeddings.

Clustering and dimensionality reduction

UMAP and spectral clustering provide the best and most intuitive results: the resulting embeddings cluster together *cuadrantes* with **high, medium, and low homicide rates.**

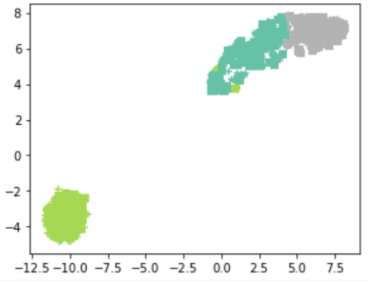


Figure: UMAP and spectral clustering over 2018-2019 homicide data.

Understanding clusters through spatial covariates

We use socio-demographic characteristics as regressors for explaining the spectral clusters created from the embeddings.

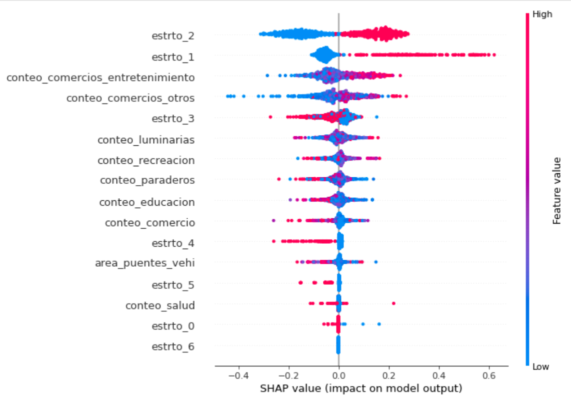
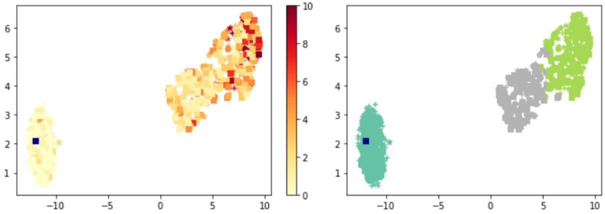


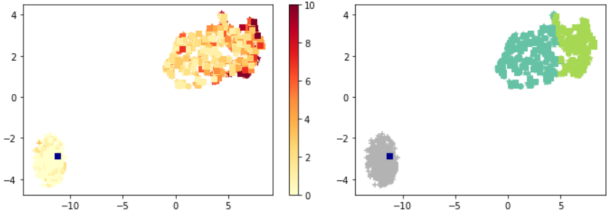
Figure: SHAP values for spatial features to explain spectral clusters.

Robustness over changing training periods

- Even though the spatial covariates explain part of the cluster assignments, these are static features that do not change over time: homicides are dynamic phenomena and therefore we expect both the embeddings and cluster assignments to vary over time.
- The overall structure of the embedding and the cluster assignments remain mostly equal. However, the clusters assignments do not remain static.

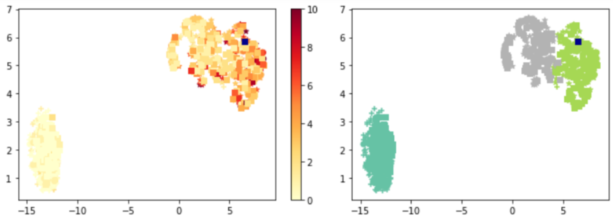


(a) Week 1

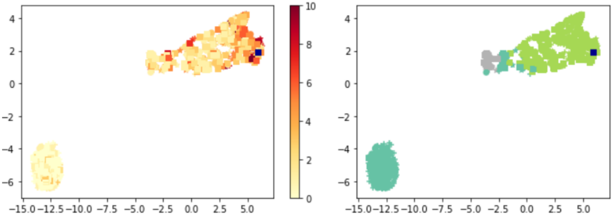


(b) Week 17

Figure: *Cuadrante* shifting from a low-rate cluster to a high-rate one over the course of a year (cluster colors may vary).



(a) Week 33



(b) Week 49

Figure: *Cuadrante* shifting from a low-rate cluster to a high-rate one over the course of a year (cluster colors may vary).

Conclusions

- Using a zero-inflated model leads to significant performance gains over the non-inflated models.
- The embeddings obtained generally have a similar structure and intuitively map the locations along a dimension according to their total homicide count.
- *Cuadrantes* are divided into three clusters, which correspond to those with high, medium, and low homicide rates, showing expected correlations with socio-demographic attributes.
- Embeddings remain generally stable over time, but they are not completely static.
- ZIE model captures relevant patterns from historic homicide data and we therefore expect to be able to use them to improve the predictive capabilities of other homicide prediction models.

Main references

- **L.-P. Liu and D. M. Blei, "Zero-inflated exponential family embeddings," in *International Conference on Machine Learning*, 2017, pp. 2140–2148.**
- G. Mohler, "Marked point process hotspot maps for homicide and gun crime prediction in chicago," *International Journal of Forecasting*, vol. 30, no. 3, pp. 491–497, 2014.
- G. Jin, Q. Wang, Z. Xia, Y. Feng, C. Qing, and H. Jincai, "Crime-gan: A context-based sequence generative network for crime forecasting with adversarial loss," *2019 IEEE International Conference on Big Data (Big Data)*, pp. 1460–1469, 2019.
- S. Zhu and Y. Xie, "Crime incidents embedding using restricted boltzmann machines," in 2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). IEEE, 2018, pp. 2376–2380.
- P. Wang, J. Zhang, G. Liu, Y. Fu, and C. Aggarwal, "Ensemble-spotting: Ranking urban vibrancy via poi embedding with multi-view spatial graphs," in Proceedings of the 2018 SIAM International Conference on Data Mining. SIAM, 2018, pp. 351–359.
- J. Yang and C. Eickhoff, "Unsupervised learning of parsimonious general-purpose embeddings for user and location modeling," ACM Transactions on Information Systems (TOIS), vol. 36, no. 3, pp. 1–33, 2018.

Graph Restrictions for Signal Processing of Homicides Data

Graph Restrictions for Signal Processing of Homicides Data

Introduction

This work was inspired by two papers and it intends to combine notions from both of them:

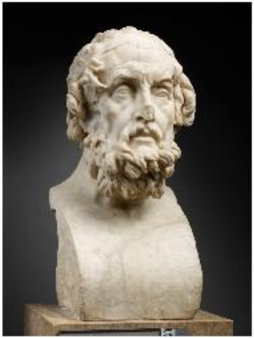
- GLoG (Homicide Prediction Using Sequential Features from Graph Signal Processing)
- Kernel Warping (A Manifold Learning Data Enrichment Methodology for Homicide Prediction)

The idea is to apply the GLoG methodology to graphs in which nodes contain any type of crime but the signal only contains homicide data.

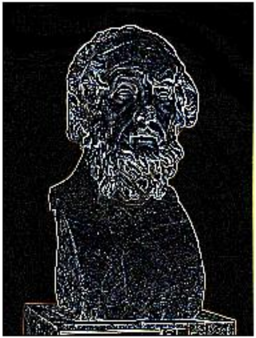
GLoG

The Laplacian of Gaussian is a methodology for edge detection commonly used in image processing.

original image



filtered image



GLoG

The Graph Laplacian of Gaussian is an extension to graphs of this methodology. It consists of 3 steps:

- 1. Graph Fourier transform.

$$GFT = U^T f \tag{1}$$

- 2. Boundary detection.

$$LoG(f) = \nabla^2 G * f \tag{2}$$

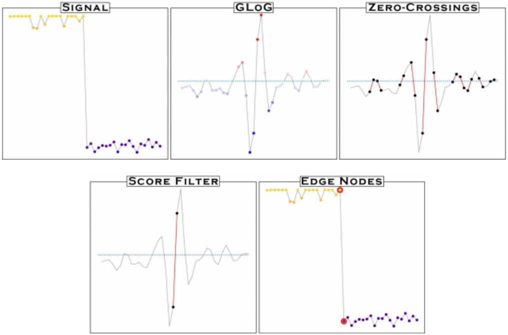
$$GLoG(f) = iGFT(\widehat{\nabla^2 G} \cdot \hat{f}) \tag{3}$$

- 3. Most relevant edge nodes.

$$GLoG(\tau_i)GLoG(\tau_j) < 0 \tag{4}$$

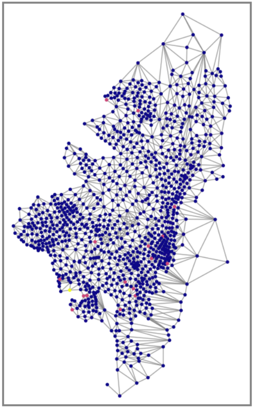
GLoG

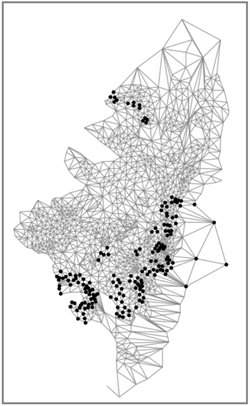
Here we see a simple example with a signal and the steps taken to find the edge nodes.



GLoG

The result of applying GLoG to a graph looks like this. On the left we have the graph with the signal and on the right the edge node configuration.





Kernel Warping

The idea of kernel warping is to construct kernels from homicide data but "warp" them to the underlying manifold with all crime data. We use Gaussian kernels.

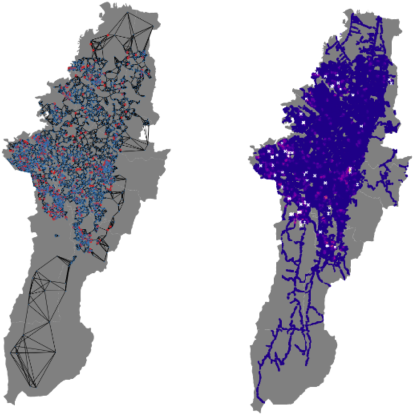
$$k(x, x_i) = \frac{1}{2\sigma^2} \text{exp} \left(-\frac{|x - x_i|^2}{2\sigma^2} \right) \tag{5}$$

Here is the warped kernel.

$$\widetilde{k}(x, s_i) = k(x, s_i) - k_{x,Z}^T \left(I + \lambda L K_{Z,Z}\right)^{-1} \lambda L k_{s_i,Z} \tag{6}$$

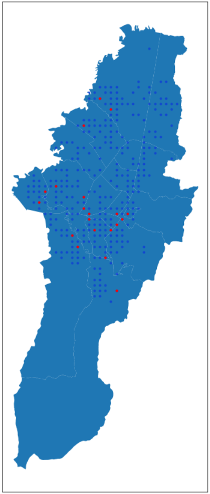
Kernel Warping

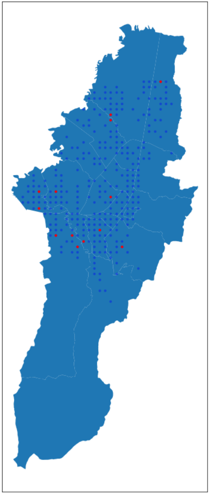
This is the warped kernel. On the left we have the crime data (blue) and homicides (red). On the right we have the intensities from the kernels and the homicides for a given month.

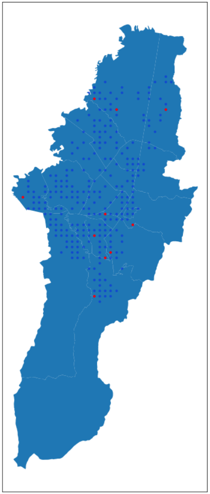


Restricted Graphs

This are the restricted graphs for 3 different weeks. The red nodes had homicides occurred in them.

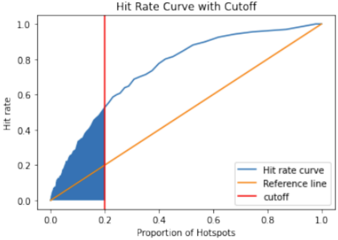




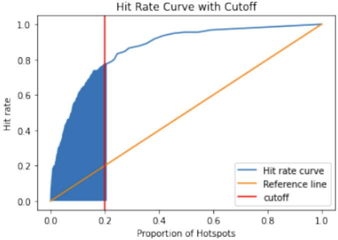


Results

The original GLoG had a HRAUC of 0.73 while the new one has 0.85. The increase is substantial but misleading. Two things changes, graph generation and graph restriction.



Original GLoG



GLoG with restricted graphs

Results

This constitutes a more just comparison. All models were trained with the same generated graph.

	ROC	Hit Rate	Hit Rate 20%
Random forest	0.831	0.838	0.092
Gradient Boosting Machine	0.842	0.852	0.089
Neural Network	0.851	0.855	0.094
Logistic Regression	0.852	0.866	0.103
Logistic Regression - Unrestricted	0.850	0.857	0.095

- The uniformly distributed grid is clearly easier to predict.
- There is a marginal gain from restricting the graph.

Dynamic Network Analysis of Spatio-Temporal Crime Incidents

Dynamic Network Analysis of Spatio-Temporal Crime Incidents

What we do...

- Basic descriptive tools of network theory to describe the topology of a spatio-temporal weekly graph of crimes in Bogotá (period 2013 to 2019).
- Centrality and community detection measures suggest time persistent structural characteristics of the graphs.
- Persistence of high page rank centrality measures suggests the existence of certain sink hotspots,
- High degree betweenness centrality of nodes suggests critical crossroads.

Introduction

- In previous papers, [?] and [?], we have explored the benefits of modeling a background graph of spatial crime incidents.
- This process has motivated us to study the graph structure described above to discover stylized facts.

Results

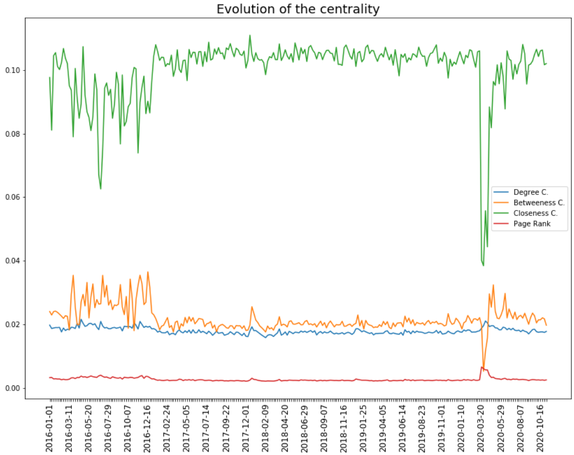


Figure: Evolution of graph centrality

Results

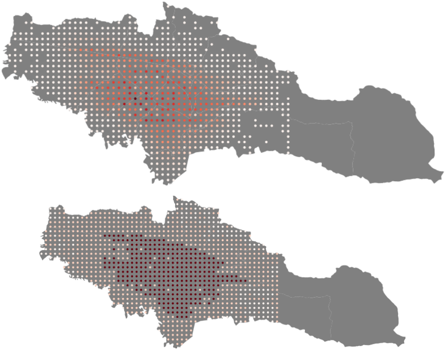


Figure: Analysis of betweenness in crime graphs.

- It is tempting to interpret the highest betweenness centrality nodes as criminal pathways among crime hotspots.
- Suggest police interventions.

Results

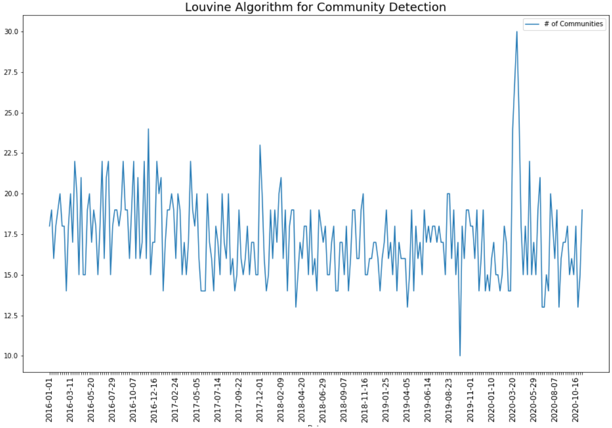


Figure: Evolution of the number of communities detected

Results

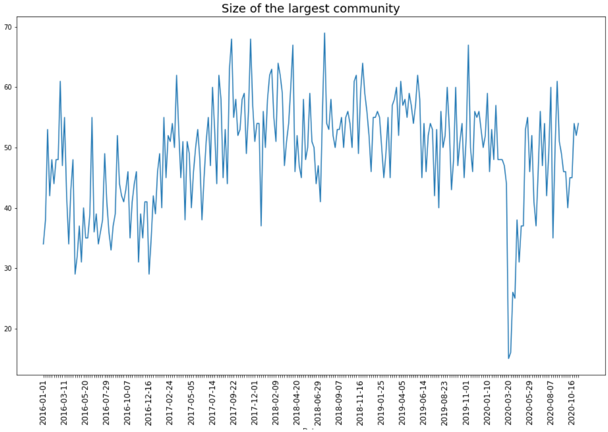


Figure: Evolution of the size of the largest community.

Results

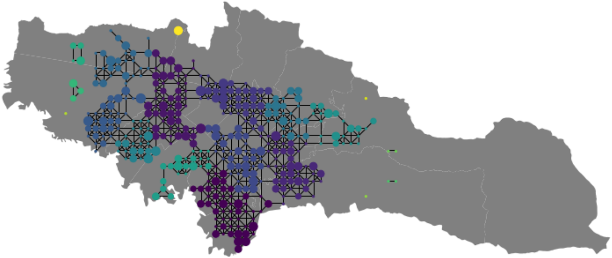


Figure: Detected communities (the size of the nodes representa the number of crimes).

Results

	Mean	Std. Deviation
Number of Nodes	384.4	58.1
Average Degree	6.8	0.8
Average Path Length	9.6	1.4
Diameter of Largest Component	27.7	4.5
Overall Clustering	0.49	0.01
Fraction of Nodes in Largest Component	0.95	0.16