<https://snap.stanford.edu/data/email-EuAll.html>

# EU email communication network

### Dataset information

The network was generated using email data from a large European research institution. For a period from October 2003 to May 2005 (18 months) we have anonymized information about all incoming and outgoing email of the research institution. For each sent or received email message we know the time, the sender and the recipient of the email. Overall we have 3,038,531 emails between 287,755 different email addresses. Note that we have a complete email graph for only 1,258 email addresses that come from the research institution. Furthermore, there are 34,203 email addresses that both sent and received email within the span of our dataset. All other email addresses are either non-existing, mistyped or spam.

Given a set of email messages, each node corresponds to an email address. We create a directed edge between nodes *i* and *j*, if *i* sent at least one message to *j*.

|  |  |
| --- | --- |
| Dataset statistics | |
| Nodes | 265214 |
| Edges | 420045 |
| Nodes in largest WCC | 224832 (0.848) |
| Edges in largest WCC | 395270 (0.941) |
| Nodes in largest SCC | 34203 (0.129) |
| Edges in largest SCC | 151930 (0.362) |
| Average clustering coefficient | 0.0671 |
| Number of triangles | 267313 |
| Fraction of closed triangles | 0.001373 |
| Diameter (longest shortest path) | 14 |
| 90-percentile effective diameter | 4.5 |

### Source (citation)

* J. Leskovec, J. Kleinberg and C. Faloutsos. [Graph Evolution: Densification and Shrinking Diameters](http://www.cs.cmu.edu/~jure/pubs/powergrowth-tkdd.pdf). ACM Transactions on Knowledge Discovery from Data (ACM TKDD), 1(1), 2007.

### Files

|  |  |
| --- | --- |
| File | Description |
| [email-EuAll.txt.gz](https://snap.stanford.edu/data/email-EuAll.txt.gz) | Email network of a large European Research Institution |