

# Gelly: Flink Graph API

Gelly 是Flink 的一种图形API，它包括一些方法(method)和工具(utility)，用来简化Flink 中图形分析应用的开发。类似批处理API，Gelly也提供一些high-level 的函数来转换(transform)、修改图形。Gelly 不仅提供创建、转换、修改图形的方法，还提供一个图形算法的库(library)。

<b>Graph API</b>
<b>Iterative Graph Processing</b>
<b>Library Methods</b>
<b>Graph Algorithms</b>
<b>Graph Generators</b>
<b>Bipartite Graphs</b>

## 使用Gelly

Gelly 现在是Maven 项目 *库* 的一部分。所有的相关类(class)都位于`org.apache.flink.graph` 包下。

要使用Gelly，在`pom.xml`里添加如下依赖。

<b>Java</b>	<b>Scala</b>
	<pre>&lt;dependency&gt;   &lt;groupId&gt;org.apache.flink&lt;/groupId&gt;   &lt;artifactId&gt;flink-gelly_2.10&lt;/artifactId&gt;   &lt;version&gt;1.3-SNAPSHOT&lt;/version&gt; &lt;/dependency&gt;</pre>

注意，Gelly 并不是二进制发行文件的一部分。将Gelly 库打包到用户的Flink 程序的方法，可参考[链接](#)。

本章的余下内容包括：对可用方法的介绍，Gelly的使用示例，以及Gelly 与Flink DataSet API混合使用方法。

## 运行Gelly 示例

[Flink distribution](#) 的`opt` 目录提供了Gelly 库的jar 文件。(低于Flink 1.2 的版本，可以从[Maven Central](#) 手动下载。) 要运行Gelly 示例，必须拷贝**flink-gelly** (for Java) 或者 **flink-gelly-scala** (for Scala) jar 到Flink 的`lib` 目录。

```
cp opt/flink-gelly_*.jar lib/
cp opt/flink-gelly-scala_*.jar lib/
```

Gelly 的示例jar 文件包含对每个库方法的驱动(driver)，可以在**examples** 目录中找到。配置完集群并启动，列出可用的算法类：

```
./bin/start-cluster.sh
./bin/flink run examples/flink-gelly-examples_*.jar
```

Gelly 驱动可以生成图形数据，或者从CSV 文件中读取边列表(集群的每个节点都必须拥有输入文件的权限)。如果选择了某个算法，算法描述、支持的输入输出、相关配置会显示出来。打印[JaccardIndex](#) 的用法：

```
./bin/flink run examples/flink-gelly-examples_*.jar --algorithm JaccardIndex
```

对有一百万个顶点(vertex)的图形，显示它的[graph metrics](#)：

```
./bin/flink run examples/flink-gelly-examples_*.jar \
  --algorithm GraphMetrics --order directed \
  --input RMatGraph --type integer --scale 20 --simplify directed \
  --output print
```

可以用 `--scale` 和 `--edge_factor` 参数调整图形的size。[library generator](#) 还提供对额外配置项的访问，用来调整幂律分布的偏度(power-law skew) 和随机噪声。

[Stanford Network Analysis Project](#) 提供了社交网络数据的样本。对入门者而言，数据集[com-lj](#) 的数据量比较适合。

通过Flink 的Web UI，运行一些算法，并监视job 的进度：

```
wget -O - http://snap.stanford.edu/data/bigdata/communities/com-lj.ungraph.txt.gz | gunzip -c
> com-lj.ungraph.txt
```

```
./bin/flink run -q examples/flink-gelly-examples_*.jar \
  --algorithm GraphMetrics --order undirected \
  --input CSV --type integer --simplify undirected --input_filename com-lj.ungraph.txt --
input_field_delimiter '$\t' \
  --output print
```

```
./bin/flink run -q examples/flink-gelly-examples_*.jar \
  --algorithm ClusteringCoefficient --order undirected \
  --input CSV --type integer --simplify undirected --input_filename com-lj.ungraph.txt --
input_field_delimiter '$\t' \
  --output hash
```

```
./bin/flink run -q examples/flink-gelly-examples_*.jar \
  --algorithm JaccardIndex \
  --input CSV --type integer --simplify undirected --input_filename com-lj.ungraph.txt --
input_field_delimiter '$\t' \
  --output hash
```

请通过用户[邮件列表](#)或者[Flink Jira](#)提交feature request，以及报告issue。我们欢迎对新算法的建议，也欢迎[贡献代码](#)。