

**تمرین سوم)** هدف از این تمرین آشنایی با الگوریتم PageRank و پیاده‌سازی آن به کمک PySpark و RDD است. شما یک گراف دارید و باید مقدار PageRank هر گره را محاسبه کنید. مقادیر این گراف در قالب یک فایل متنی (Graph\_Links.txt) به شما داده شده است. هر خط از فایل شامل شماره یک گره و لیستی از گره‌هایی است که از گره به آن‌ها لینک داده شده است.

(۱) داده‌های موجود در فایل متنی را بارگذاری کرده و گراف مربوطه را رسم کنید.

(۲) مشخص کنید کدام گره‌ها بصورت Dead End هستند.

(۳) الگوریتم PageRank را برای گراف مربوطه پیاده‌سازی و اجرا کرده و مقادیر PageRank تمامی گره‌ها را ذخیره کنید. ( $\beta =$

0.8 و معیار همگرایی کمتر از 0.001 فرض شود).

(۴) الگوریتم HITS را پیاده‌سازی کنید. برای نرمال‌سازی، بیشترین امتیاز هر کدام از بردارهای Hubs و Authorities را برابر با یک

قرار دهید. همچنین بروزرسانی‌ها را انقدر تکرار کنید تا به همگرایی برسید. حال مقادیر Hubs و Authorities تمامی گره‌ها را

ذخیره کنید.