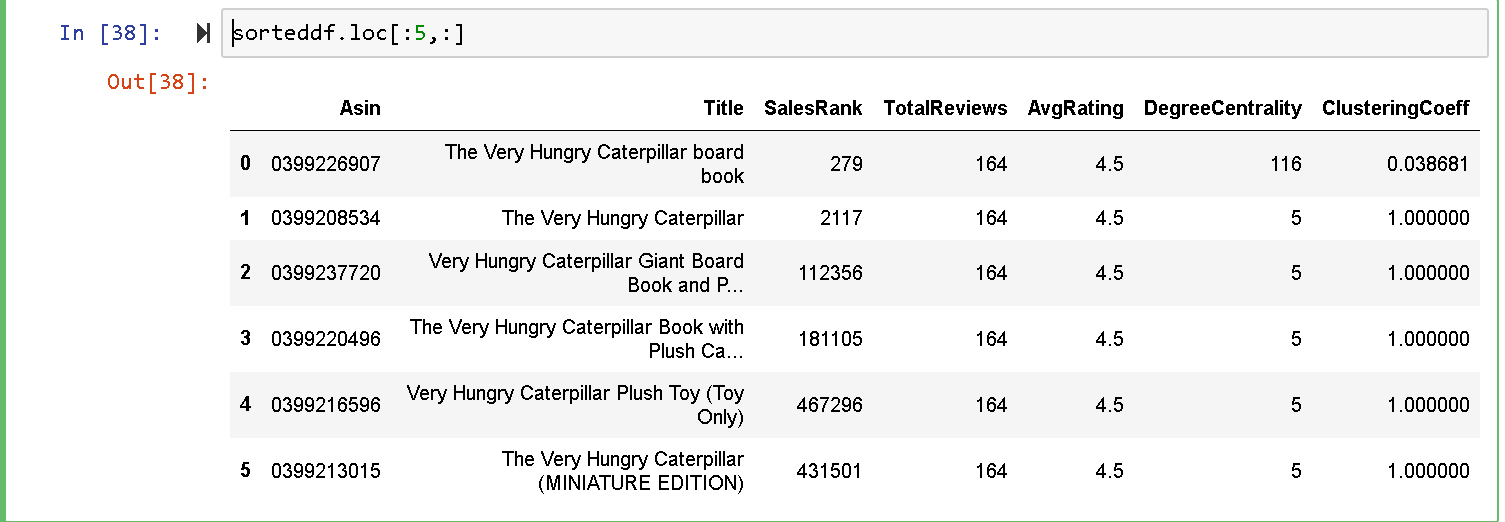
**Social Network Analysis – Assignment-3**

**Output Results:**



**Logic Behind 4th step:**

Composite score is calculated using the formula:

**df["composite measure"] = df["Weight"]\*(df["exp\_TotalReviews"]\*df["AvgRating"])/(df["mms\_SalesRank"]+1)**

1. Df[“Weight”] = the edge weight which determines Category similarity, if the edge weight is more then, there exists high similarity between this node and the node of interest.

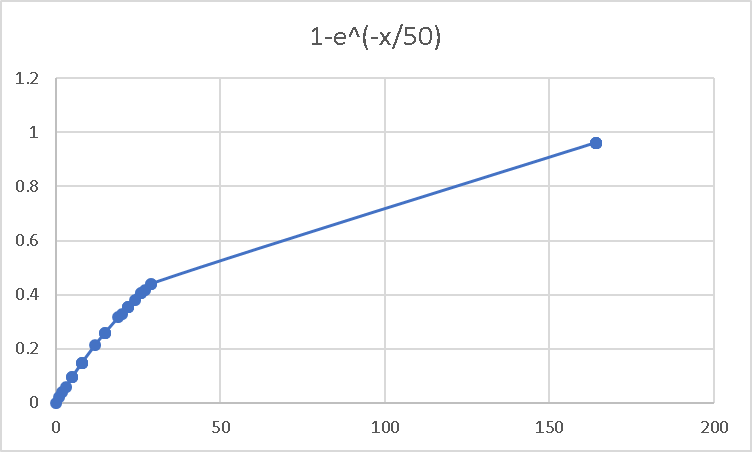
This is one of the important features to consider as this drives us to the interest of the node.

1. **Df[“exp\_TotalReviews”] = 1-np.exp(-df["TotalReviews"]/50) = 1-e^(-x/50)**

Here, the values of TotalReviews ranges from 0 to 164. This feature is important as we can only recommend a book to someone if it is reviewed.

**If TotalReviews = 0** then 1-e^0 = 0. When this is multiplied in the equation then composite score becomes 0, which tells us that we cannot recommend this book as it is not reviewed by anyone.

**Divided by 50:** This is the tolerance limit which we can set to any number. Here I have chosen 50 by plotting different graphs. We scale this so that the one with more reviews dominate and the one with less reviews will be supressed. For suppose If our tolerance is 10 then the curve looks like in the second graph, in the 2nd graph after certain point on x axis the curve gives the output close to 1 i.e., most of them, even if the reviews are less than 50 will have equal importance as the values above 150.



1. df[“AvgRating”] = this value ranges from 4 to 5 for the given purchaseAsin. For any uknown Asin the value ranges from 0 to 5 multiplying this value directly with the scaled value of total reviews and edge weights improves the value in the numerator.
2. df["mms\_SalesRank"] + 1 = MINMAX scaling the SalesRank removes all the outliers and helps in scaling the values. Since it is mentioned as low sales rank is the best-selling book, if we divide our composite measure with scaled values then, the highest value obtained will be our recommendation to the node of interest.
3. Here we add +1 in the denominator as the values of the SalesRank after using MINMAX scalar ranges from 0 to 1. Adding 1 eliminates 0 in the Denominator and scales the values to range from 1 to 2