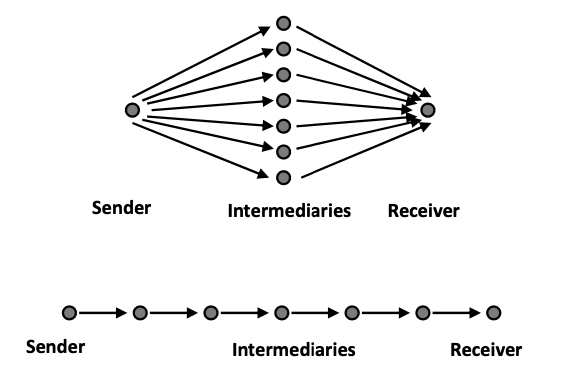
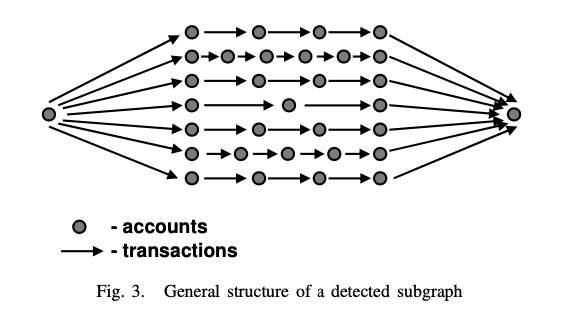
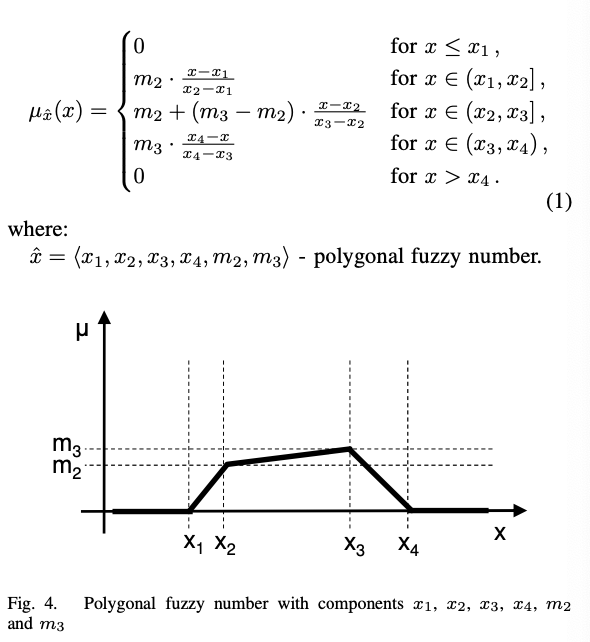
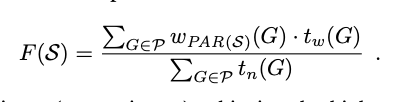
Datos importantes de: Graph Mining Approach to Suspicious Transaction Detection – Krzysztof Michalak, Jerzy Korczak.

* Las técnicas de machine learning pueden ser no muy eficientes al ser aplicadas a la detección de transacciones sospechosas.
* Este documento propone una nueva metodología de machine learning para minar gráficas de transacciones.
* Dentro de todas las transacciones que hacen los bancos, los casos fraudulentos son muy pocos, en los ejemplos que da el artículo están entre 0.02-0.04%. Esto hace que sea muy difícil de aplicar métodos de machine learning.
* El proceso de lavado de dinero se divide en tres etapas: colocación, poner en capas e integración. Cada una de estas etapas involucra varios esquemas de transferencia de dinero entre varias cuentas. Algunos ejemplos son estos:
* El primer caso es para dividir el dinero y que las transacciones no sean reportadas como sospechosas.
* El segundo caso es para que se pierda el hilo de cuál es la cuenta que manda el dinero.
* A veces algunas de las transacciones intermediarias sí son de actividades legales.
* La minería de datos es interesante porque puede encontrar dependencias complejas de transacciones, es decir, les sigue el hilo a las transferencias para ver si hay algo de este tipo:
* Para modelar estos subgrafos, se genera un **patrón jerárquico de tres niveles**.
  + TR=<a,r,s>
    - a-cantidad de la transacción.
    - s, r-función que le asigna peso a las clases que las entidades mandan y reciben (respectivamente) pertenecen.
  + SER=<m,delta>
    - m-cantidad de transacciones en la cadena.
    - delta-radio que representa la cantidad que se mandó en la última transacción y en la primera transacción.
  + PAR=<n, DELTA, tau>
    - n-número de transacciones en la cadena.
    - DELTA-radio que representa la suma de las cantidades recibidas por la última cuenta y la suma que manda la cuenta primera.
    - tau-umbral de aceptación que se usa para decidir cuáles transacciones forman parte del patrón.
* En el método propuesto se usan **simplified polygonal fuzzy numbers**.
  + Los números borrosos son un concepto que se usa en teoría de conjuntos difusos (es un conjunto en la que cada elemento tiene un grado asociado entre 0 y 1 de qué tan probable es que ese número forme parte de ese conjunto, en lugar de tener conjuntos en los que o perteneces o no perteneces) los simplified polygonal fuzzy numbers pueden considerarse como una generalización de los números reales a los conjuntos borrosos.
* En la imagen se muestra un número borroso poligonal donde mi es qué tan probable es que el número xi forme parte de este conjunto.
* A cada una de las transacciones se les asigna un peso denotado como wTR=mu\*s\*r
  + Donde mu es la probabilidad de que a sea parte del conjunto.
* Para entrenar al algoritmo se necesitan las anotaciones de un experto.
* Se declaran muchas más variables y al final se llega a que la evaluación de un espécimen está calculado por la siguiente función:
  + 
* Los experimentos del artículo fueron hechos con datos generados artificialmente.
* La generación de esquemas de lavado de dinero se caracteriza por las siguientes funciones de distribución:
  + La distribución de la cantidad mandada por el nodo inicial al intermediario.
  + Distribución del número de intermediarios.
  + Distribución de la fracción de lo que recive el intermediario y lo que recibe el último nodo.
* Después se dan datos sobre los experimentos que se hicieron y se detiene mucho a explicar cómo construyeron cada uno de los conjuntos de datos.
* Algo que no me queda muy claro es si evalúa nodo por nodo para poder construir el subgrafo y si guarda una variable que representa todos los nodos anteriores al que se está evaluando.
* Podría complementarse con la tesis que vimos porque la tesis se centra más en la forma de optimizar la tarea de tirar la red, más que en detectarla. Este artículo se centra en la detección de la red y si se entiende bien el algoritmo propuesto, sí podría ser aplicable.

El objetivo es hacer un documento. Trabajar en un documento reproducible, como un Markdown. Ir poniendo los datos del resumen. Aplicar el análisis que propone el paper a los datos simulados.

Voy a hacer un Markdown de un análisis descriptivo de datos, subitulo de teoría de redes, aplicación (o al revés). Pego los datos de las tres tablas. Tener certeza de que los datos se vean bien. A ver si se puede hacer gráficas de la red o ver qué se podría sacar de la red. Dentro de tres semanas ir viendo isn podemos identificar SAR.

Sobre estos datos, pasar criterios de identificación de anomalías, no son específicos para redes, pero ver si identifican alguna operación rara.