

DETECCIÓN DE COMUNIDADES EN TOPOLOGÍAS DE REDES ELÉCTRICAS: EFECTO DE LA INCERTIDUMBRE DE LOS PESOS DE LOS ENLACES

JOSÉ A. MORONTA R. ¹, CLAUDIO M. ROCCO S. ²

¹ Universidad Simón Bolívar, Departamento de Tecnología Industrial. e-mail: jmoronta@usb.ve

² Universidad Central de Venezuela, Facultad de Ingeniería. e-mail: croccouv@gmail.com

Recibido: febrero 2016

Aprobado para publicación: junio 2016

RESUMEN

Una comunidad o estructura modular puede ser definida como un sub conjunto de nodos de una red que están más densamente conectados entre sí que con el resto. La detección de las comunidades en una red depende de la topología de la misma, así como de las características de los enlaces que unen a dichos nodos (por ejemplo: longitud, capacidad o confiabilidad). Este trabajo propone evaluar cómo la incertidumbre de los pesos afecta la detección de comunidades. La propagación de la incertidumbre es evaluada mediante el uso de simulación de Monte Carlo. A partir de la topología de la red seleccionada con pesos fijos, se determina la partición base. A medida que se generan valores aleatorios de los pesos, de acuerdo con distribuciones previamente seleccionadas, se determinan nuevas particiones que son comparadas con la partición base a través del índice *variation information* (v_i), una métrica que permite evaluar la similitud entre particiones. El análisis de la aproximación de la distribución asociada a v_i , permite evaluar los efectos de la incertidumbre. La propuesta se ilustra sobre una red de prueba y la topología de dos sistemas eléctricos de potencia reales. Los resultados indican que la detección de comunidades en sistemas eléctricos es sensible a la variación del peso de los enlaces y la magnitud de los efectos depende del tamaño de la perturbación así como de la topología de la red en estudio. Los autores consideran que la información obtenida mediante el enfoque propuesto, aunado a un posterior análisis de sensibilidad, puede servir para establecer límites en la variación estructural de las comunidades. Este análisis es importante en aquellos estudios que utilizan como insumo la partición de la red.

Palabras clave: detección de comunidades, Monte Carlo, sistemas de potencia, propagación de incertidumbre

COMMUNITY DETECTION IN ELECTRIC NETWORK TOPOLOGIES: EFFECT OF UNCERTAINTY OF THE WEIGHT OF LINKS

ABSTRACT

In this paper, the propagation of uncertainty in the detection of communities in networks with weighted links is assessed. Communities can be defined as subsets of nodes which are densely connected to each other and loosely connected to the rest of the network. The weights represent particular characteristics associated to the links of a real network (e.g., length, capacity or reliability). The uncertainty of the weights of links is modeled using known probability functions. The effects of such uncertainty on the structure of communities are evaluated by using a Monte Carlo approach: each partition generated during a sample is compared to a previously defined base case partition using the similarity index *variation information* (v_i). The approximated probability distribution of v_i is used for quantifying the effects of uncertainties. The proposed approach is illustrated using a benchmark network and the topology of two real electrical power systems. The results show that the detection of communities in electrical systems is sensitive to the variation of the weight of the links and that the magnitude of the effects depends on the size of perturbation analyzed as well as on the topology of the network. The authors consider that the information obtained through the proposed approach and a posteriori sensitivity analysis can be used to set limits on the structural variation of the base communities. This analysis is very important on studies that use the network partition as input to other studies.

Keywords: community detection, Monte Carlo, power systems, uncertainty propagation

INTRODUCCIÓN

En sistemas susceptibles a ser modelados como redes o grafos es posible encontrar “estructuras modulares” o “comunidades” definidas como un conjunto de nodos que están más densamente conectados entre ellos que con el resto de la red. Una característica de las estructuras modulares es que los nodos pertenecientes a una misma comunidad comparten características comunes, atributos o funcionalidad similar (Fortunato, 2010). La detección de comunidades o estructuras modulares en redes, se ha venido aplicando en diversos campos desde las redes sociales (Zachary, 1977; Girvan et al., 2002 y Clauset et al., 2004), redes de transporte, servicios e intercambio de bienes (Clauset et al., 2004; Boccaletti et al., 2006 y Rocco et al. 2011).

El concepto de estructura modular o comunidad no está definido rigurosamente, y depende de cierto grado de arbitrariedad (Fortunato, 2010), incluso del ámbito específico de aplicación. Por esto, el proceso de definición de la cantidad de grupos de nodos (comunidades) y la ubicación de nodos en éstos, depende de: el algoritmo con que se realice la partición, la topología de la red, las características de los elementos de la red, entre otros aspectos.

Los enlaces de una red representan la interacción entre los nodos, y ésta se cuantifica asociando un peso o ponderación a estos enlaces. Según la red en estudio, este peso puede representar, por ejemplo, número de vuelos por semanas (Bagler, 2008), potencia eléctrica reactiva transmitida (Mei et al., 2011), probabilidad de falla, entre otras características.

Para las redes con enlaces ponderados se han desarrollado algoritmos particulares de detección de comunidades. Algunos autores han cuantificado la fortaleza de los algoritmos de detección de comunidades en redes ponderadas: Fan et al. (2006) estudian el efecto de los pesos de los enlaces y concluyen que “la redistribución de pesos afecta fuertemente la estructura comunitaria de las redes”. En un trabajo posterior Fan et al. (2007) investigan la exactitud y precisión de algunos algoritmos de detección en redes ponderadas, sin ponderar y usando ponderación inversa.

Por otra parte, Karner et al. (2007) consideran sólo variaciones en la topología de las redes e indican “que la importancia de la estructura comunitaria se puede cuantificar de manera efectiva mediante la medición de su robustez frente a pequeñas perturbaciones en la estructura de la red”. La evaluación de las particiones se realiza

mediante índices de similitud. Estos índices normalizados en $[0,1]$ permiten definir si dos particiones obtenidas son iguales o diferentes, dejando al lector determinar, en forma cualitativa, si valores intermedios corresponden a particiones iguales o no.

En este trabajo se presenta una propuesta para: 1) evaluar el efecto de la incertidumbre de los pesos de los enlaces en la detección de comunidades y 2) aproximar la distribución de probabilidad del índice de similitud de particiones seleccionado. Esta propuesta se ilustra en una red de prueba de la literatura, (Newman et al., 2004), y en la topología de dos sistemas eléctricos reales.

La estructura del trabajo es como sigue: la sección detección de comunidades presenta algunas definiciones fundamentales asociadas a la detección de comunidades; la sección comparación de particiones describe cómo se puede realizar la comparación entre particiones distintas; la sección procedimiento describe los pasos necesarios para la determinación de la incertidumbre asociada al índice de similitud; la sección resultados analiza los resultados obtenidos. Finalmente se presentan las conclusiones y trabajos futuros.

DETECCIÓN DE COMUNIDADES EN REDES PONDERADAS

A. Grafo ponderado

Un grafo o red ponderada $G^W=(V,E,W)$ consiste en un conjunto $V=\{v_1, v_2, \dots, v_m\}$ de nodos (o vértices), un conjunto $E=\{e_1, e_2, \dots, e_n\}$ de enlaces que conectan los vértices, y un conjunto de números reales (pesos) $W=\{w_1, w_2, \dots, w_n\}$ asociados a los enlaces y que representan alguna característica particular de los mismos (por ejemplo, impedancia, confiabilidad).

B. Comunidad o estructura modular

Dado un grafo $G^W = (V,E,W)$, una comunidad es un subgrafo $G'^W = (V',E',W')$, cuyos nodos presentan una fuerte cohesión estructural, que puede ser cuantificada de distintas maneras (Boccaletti, 2006).

C. Partición

Una partición C es la división de una red o grafo en comunidades, considerando que cada vértice o nodo del grafo pertenecen a alguna comunidad. Al incrementarse el número de nodos de una red crecen las posibles particiones de forma exponencial (Aldecoa, 2012). Por esto, es

fundamental definir la partición óptima para una red. Sin embargo, la determinación de la partición óptima está asociada a la definición de indicadores de la “calidad” de la partición seleccionada.

D. Partición base

Es la partición obtenida con los valores nominales o base de los pesos de los enlaces.

E. Modularidad

La modularidad es la medida más empleada “para determinar la fortaleza de las comunidades encontradas por los métodos de detección de comunidades” (Newman, 2006).

La modularidad Q mide la fracción de los enlaces de la red que conectan los nodos del mismo tipo (es decir, los enlaces dentro de la comunidad) menos el valor esperado de la misma cantidad en una red con igual número de comunidades, pero considerando conexiones al azar entre los nodos. Los valores que se acercan al valor máximo, $Q=1$, indican la posible presencia de estructuras modulares. En la práctica, los valores de modularidad se encuentran en el intervalo 0,3 a 0,7 (Newman, 2006).

En Newman (2004) se propone una expresión para cálculo de la modularidad en redes ponderadas:

$$Q^w = \frac{1}{2 * m} \sum_{ij} \left[P_{ij} - \frac{p_i p_j}{2m} \right] \delta(c_i, c_j) \quad (1)$$

donde:

m : la suma de todos los pesos de los enlaces de la red

P_{ij} : el peso de los enlaces entre los nodos i y j

p_i, p_j : los pesos de los nodos i y j .

$$p_i = \sum_j P_{ij} \quad (2)$$

$\delta(c_i, c_j)$: es 1 si $c_i = c_j$, y 0 en cualquier otro caso.

c_i : Comunidad a la cual es asignado el nodo i

c_j : Comunidad a la cual es asignado el nodo j

F. Algoritmos de detección de comunidades

Se han desarrollado diversos algoritmos para la detección de comunidades en grafos, cada uno con sus especificidades en cuanto a: complejidad computacional, medidas para la validación de las comunidades, entre otros indicadores (Fortunato, 2010). Este trabajo no está enfocado a la selección del mejor algoritmo; por ello sólo se ilustra el

algoritmo *Fastgreedy*, el cual puede ser utilizado sobre redes ponderadas o no. Adicionalmente, el algoritmo seleccionado fue considerado como idóneo para la detección de comunidades en sistemas eléctricos de potencia, en la comparación realizada en (Moronta y Rocco, 2016).

Básicamente el algoritmo busca determinar los cambios en la modularidad que se derivan a partir de la fusión iterativa de comunidades (Clauset et al., 2004). Las comunidades que el algoritmo detecta son excluyentes, es decir los nodos sólo pertenecen a una única comunidad.

El algoritmo *Fastgreedy* se describe como:

1. Inicialmente, cada nodo de la red es considerado como una sola comunidad;
2. Calcular el cambio en la modularidad del grafo (ΔQ_{ij}) al crear una nueva comunidad con las comunidades i y j ;
3. Combinar las dos comunidades que generan el mayor cambio en la modularidad (ΔQ_{ij});
4. Repetir los pasos 2 y 3 hasta que el cambio de la modularidad sea negativo ($\Delta Q_{ij} \leq 0$).

El resultado de la aplicación del algoritmo de detección de comunidades produce como resultado:

a) Una partición de la red o grafo:

- Agrupaciones de nodos, elementos del conjunto V de nodos, que conforman las comunidades.
- Enlaces intercomunidades, elementos del conjunto E de enlaces, que conectan las comunidades.

b) Un valor de modularidad.

Para ilustrar el algoritmo *Fastgreedy* sobre redes ponderadas, se considera la red mostrada en la Figura 1 (Rocco et al., 2011) y los valores de los pesos asociados a cada enlace.

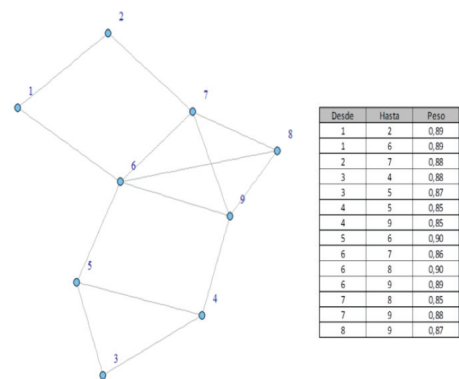


Figura 1. Red ponderada para ilustrar el algoritmo *Fastgreedy*

La Tabla 1 muestra los resultados de las iteraciones de agrupación de nodos hasta maximizar la modularidad (iteración 7) y obtener la partición mostrada en la Figura 2.

Tabla 1. Resultados de las iteraciones del algoritmo Fastgreedy para la red de la Figura 1

Iteración	Comunidades combinadas		Comunidades	ΔQ^m	Q^m
0	---		{1},{2},{3},{4}, {5},{6},{7},{8},{9}	-0,1226	---
1	{1}	{2}	{1,2},{3},{4},{5}, {6},{7},{8},{9}	-0,0604	0,0622
2	{3}	{4}	{1,2},{3,4},{5}, {6},{7},{8},{9}	-0,0037	0,0567
3	{5}	{3,4}	{1,2},{3,4,5}, {6},{7},{8},{9}	0,0989	0,1026
4	{8}	{9}	{1,2},{3,4,5}, {6},{7},{8,9}	0,1394	0,0405
5	{7}	{8,9}	{1,2},{3,4,5}, {6},{7,8,9}	0,2100	0,0706
6	{6}	{7,8,9}	{1,2},{3,4,5}, {6,7,8,9}	0,2847	0,0747
7	{1,2}	{6,7,8,9}	{1,2,6,7,8,9}, {3,4,5}	0,2635	-0,0212

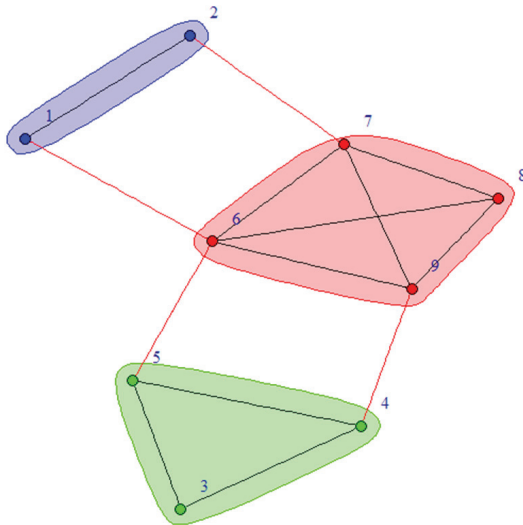


Figura 2. Partición de la red de la Figura 1 mediante el algoritmo Fastgreedy

COMPARACIÓN DE PARTICIONES

El uso de diferentes algoritmos y/o técnicas para la detección de comunidades en redes, así como la comparación de particiones obtenidas a partir de la variación de alguna de las características de la red original, genera la necesidad de emplear medidas de comparación para establecer la

similitud entre los diversos resultados obtenidos. En el caso de este trabajo, la variación del peso de los enlaces es la variable que posiblemente origine la diversidad de particiones en una misma red.

Dentro de las medidas para determinar la similitud de dos particiones, destacan las siguientes: *Rand Index* (Rand, 1971), *Adjusted Rand Index* (ARI) (Hubert et al., 1985), *Variation Information* (vi) (Meila, 2003) y *Normalized Mutual Information* (Danon et al., 2005).

Por lo general, las medidas se normalizan en el intervalo [0,1]: uno de los extremos significa que las particiones comparadas son independientes o totalmente distintas y el otro que las particiones comparadas son idénticas. Por ejemplo, para el *Rand Index*, el valor extremo uno indica que las particiones son idénticas, mientras que para la medida *Variation Information*, el extremo cero identifica la igualdad de las particiones.

Sin embargo, en la literatura no existe un valor crítico o de referencia que permita concluir cuándo dos particiones pueden ser consideradas idénticas o totalmente distintas. En general, algunos autores como Hric et al. (2014), Marotta et al. (2014), han establecido valores cualitativos, pero sin ninguna base teórica. Por ejemplo, Marotta et al.(2014) sugieren que valores del índice ARI superiores o iguales a 0,8 indican particiones iguales. De aquí la necesidad de considerar el comportamiento estadístico del índice de similitud.

En este trabajo se emplea el índice *Variation Information*, ya que éste, al representar una distancia, permite asociar más fácilmente el concepto de similitud (Meilă, 2003). Sin embargo, cualquier índice de los mencionados pudiera ser utilizado.

Variation Information (vi)

La comparación de particiones por el método de *Variation Information* (Meilă, 2003) utiliza la probabilidad de que un nodo seleccionado pertenezca a una comunidad particular.

Sea C la partición base, y C' la partición con la cual se compara. Si la selección de los nodos se supone equiprobable, entonces la probabilidad de seleccionar un nodo de una red de n nodos y que éste pertenezca a la comunidad C_k es:

$$P(k) = \frac{n_k}{n} \quad (3)$$

donde:

n_k = número de nodos en la comunidad k .
 n = número total de nodos de la red.

Sea $P(k)$ una variable aleatoria que toma K valores, donde K es el número de comunidades de la partición C . La incertidumbre asociada a seleccionar un nodo de la red de n nodos y que éste pertenezca a la comunidad C_k , se define como la entropía de la variable aleatoria $P(k)$:

$$H(C) = -\sum_{k=1}^K P(k) * \log P(k) \quad (4)$$

Sea $P(k, k')$ la probabilidad de que un nodo pertenezca a C_k en la partición C y a $C_{k'}$ en C' , es decir:

$$P(k, k') = \frac{|C_k \cap C_{k'}|}{n} \quad (5)$$

Sea $I(C, C')$ la información mutua entre las particiones C y C' definida como:

$$I(C, C') = \sum_{k=1}^K \sum_{k'=1}^{K'} P(k, k') * \log \frac{P(k, k')}{P(k)P(k')} \quad (6)$$

El índice vi de comparación entre dos particiones se define como:

$$VI(C, C') = H(C) + H(C') - 2I(C, C') \quad (7)$$

Esta medida es un valor no negativo, y es igual a cero si las particiones comparadas son exactamente iguales. El valor máximo es $\ln(n)$, donde n es el número total de nodos en la red, y ocurre si se comparan dos particiones totalmente diferentes. Para realizar la comparación entre redes distintas, es necesario normalizar el valor de vi dividiéndolo entre $\ln(n)$.

En la Tabla 2, se muestra el valor del índice vi al comparar la partición de la Figura 2 (última partición obtenida mostrada en la Tabla 1) con las particiones obtenidas en los pasos anteriores.

Tabla 2. Resultados de la comparación de particiones mediante el índice vi

Particiones comparadas		Índice vi
{1,2},{3,4,5}, {6,7,8,9}	{1},{2},{3},{4},{5},{6},{7}, {8},{9}	1,13636
{1,2},{3,4,5}, {6,7,8,9}	{1,2},{3},{4},{5},{6},{7},{8},{9}	0,98233
{1,2},{3,4,5}, {6,7,8,9}	{1,2},{3,4},{5},{6},{7},{8},{9}	0,82830
{1,2},{3,4,5}, {6,7,8,9}	{1,2},{3,4,5},{6},{7},{8},{9}	0,61613
{1,2},{3,4,5}, {6,7,8,9}	{1,2},{3,4,5},{6},{7},{8,9}	0,46210
{1,2},{3,4,5}, {6,7,8,9}	{1,2},{3,4,5},{6},{7,8,9}	0,24993
{1,2},{3,4,5}, {6,7,8,9}	{1,2},{3,4,5},{6,7,8,9}	0

El valor más alto para el índice vi (1,13636), corresponde a la comparación de la partición final con la red original (es decir, cada nodo es considerado una comunidad). En los pasos posteriores, a medida que las particiones se asemejan a la partición final, el valor de vi va disminuyendo hasta obtener el valor 0, que indica que las dos particiones comparadas son idénticas.

PROPAGACIÓN DE INCERTIDUMBRE

La propagación de incertidumbre se define como la cuantificación de la incertidumbre a la(s) salida(s) de un modelo, debido a la incertidumbre en la(s) entradas(s). En este trabajo, las entradas son los pesos de la red en estudio y la incertidumbre es modelada a través de variables aleatorias con distribución conocida. La salida corresponde a uno de los índices de similitud mencionados anteriormente.

La técnica seleccionada para evaluar la propagación de incertidumbre es la técnica de simulación de Monte Carlo (Klindt, 2000). Dada una función de salida $y=g(x)$, donde $x=(x_1, x_2, x_3, \dots, x_n)^t$ es un vector de variables aleatorias con función de distribución de probabilidad conocida $f_i(x_i)$, la técnica consiste en:

- 1) Generar n desviaciones aleatorias independientes para cada variable x_i ;
- 2) Evaluar $g(x)$ en el vector generado en 1);
- 3) Repetir 1) y 2), ns veces.

El conjunto de ns evaluaciones de $g()$ constituye una aproximación gradual de la función de densidad de

probabilidad de la salida, en la que se pueden calcular diversas estadísticas (por ejemplo, percentiles, desviaciones, entre otras (Pesonen et al., 1996)).

PROCEDIMIENTO

Incertidumbre de los pesos de los enlaces en la detección de comunidades

A partir del grafo asociado a la red bajo estudio y las distribuciones de los pesos de los enlaces del grafo:

1. Se determina la partición base de la red ponderada (a partir de los valores iniciales de W_0), mediante el método de detección de comunidades *Fastgreedy*.
2. A partir de las distribuciones asociadas a los pesos de los enlaces se generan ns vectores de muestras.
3. Con estas muestras, se determinan las particiones de la red ponderada con los valores generados en 2, mediante el método de detección de comunidades *Fastgreedy*.
4. Se comparan, mediante el índice de similitud seleccionado (en este trabajo se emplea el índice *Variation Information*), cada una de las particiones obtenidas en 3 con la partición base (obtenida en 1), generando así un vector de tamaño ns que permite determinar una aproximación a la distribución de probabilidades del índice seleccionado.

Para efectos ilustrativos se supone:

- a) Los pesos W_0 asociados al caso base se generan a partir de una distribución uniforme $U[0,85;0,90]$.
- b) La incertidumbre en los pesos se modela a través de una distribución uniforme. Se analizan tres casos: a) $U[W_{ij} \pm 0,01 * W_{ij}]$; b) $U[W_{ij} \pm 0,05 * W_{ij}]$; y c) $U[W_{ij} \pm 0,1 * W_{ij}]$, que corresponden a variaciones de 1; 5 y 10 % de los valores bases.
- c) La topología es conocida y no se consideran variaciones en la topología de la red, durante el procedimiento descrito.
- d) Se generan $ns=10000$ muestras.

La detección de comunidades, los algoritmos para la simulación de Monte Carlo y la evaluación del índice de comparación se realizaron mediante algoritmos implementados por los autores en el software libre R (R Project, 2014), versión 3.1.1, plataforma: i386-pc-mingw32/i386 (32-bit), y la librería igraph 0.7.1 y ejecutados en un computador Intel® Pentium® Dual CPU T3200 @ 2.00 GHz, 2.00 GB de RAM.

Redes seleccionadas

A.1 Red de prueba

Esta red fue propuesta por (Newman et al., 2004) y ha sido empleada por diversos autores como red de comparación (*benchmark*). La red está formada por 128 nodos con cuatro comunidades claramente definidas. La red se genera de manera artificial asignando los nodos a las comunidades, y generando interconexiones internas y con las otras comunidades con probabilidades preestablecidas. En la Figura 3 se muestra la partición base para esta red identificando claramente las cuatro comunidades resultantes. Los enlaces en color negro identifican los enlaces intra-comunidades mientras que los enlaces en color rojo identifican los enlaces inter comunidades. Los nodos que pertenecen a una misma comunidad se identifican mediante colores iguales.

A.2 Sistema eléctrico venezolano (Red VEN)

Es una red de 635 nodos y 803 enlaces (Rocco, 2008), que representa el sistema eléctrico venezolano en un año en particular, desde el nivel de tensión de 115kV hasta 765kV. En la Figura 4 se muestra la partición base para esta red (por razones de confidencialidad, la topología se muestra sin ninguna referencia geográfica). Cada área sombreada representa a una comunidad, para un total de 21.

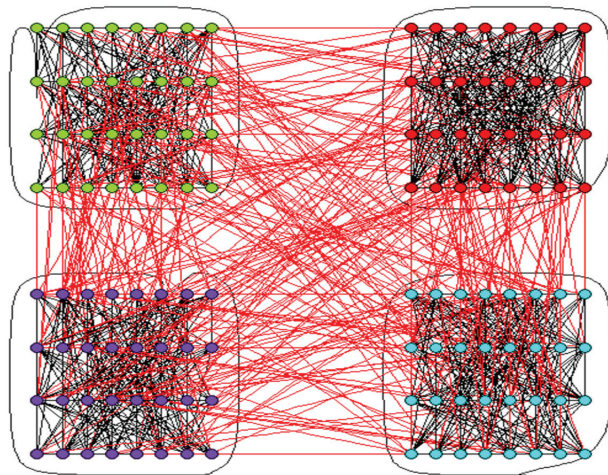


Figura 3. Partición base de la red benchmark

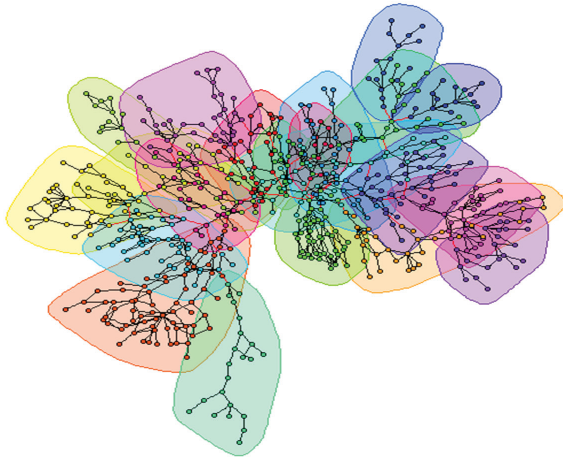


Figura 4. Partición base del Sistema Eléctrico Venezolano (Red VEN)

A.3 Sistema IEEE 118 (Red IEEE)

Es una red de pruebas de la IEEE, compuesta por: 186 enlaces y 118 nodos (Power System Test Case, 2015). Esta red representa una parte del sistema eléctrico norteamericano (específicamente en el medio Oeste de los Estados Unidos), en diciembre de 1962. En la Figura 5, se muestra la partición base para esta red, compuesta por ocho comunidades.

RESULTADOS

A.1 Red artificial

Esta red artificial no presenta mayores variaciones en el índice de similitud normalizado (SI) como puede observarse en la Figura 6, la cual representa una aproximación a la distribución acumulada de probabilidades, para cada uno de los porcentajes de incertidumbre de los pesos de los enlaces.

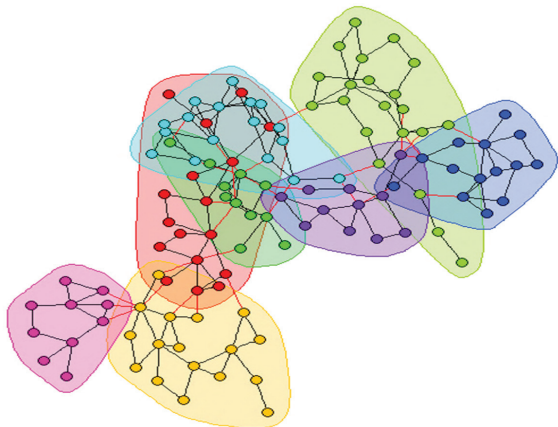


Figura 5. Partición base del Sistema IEEE de 118 nodos (Red IEEE)

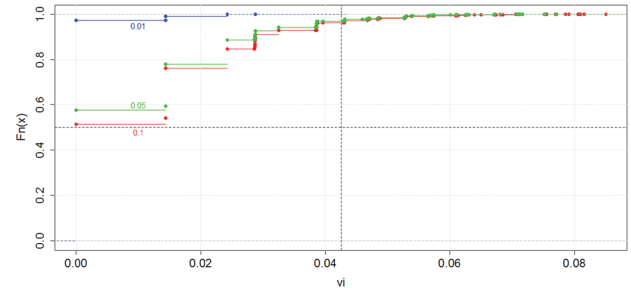


Figura 6. Función de distribución de probabilidad acumulada para el índice v_i de la red de prueba, para diferentes ΔW_i

Si se supone que un valor de $v_i \leq 0,1$ indica la presencia de particiones iguales, tenemos que, inclusive para la mayor desviación de la perturbación ($\pm 10\%$), la probabilidad de que las particiones sean idénticas es prácticamente 1. Esto indica que para esta red, las variaciones en los pesos de los enlaces no afectan las comunidades detectadas en el caso base.

A.2 Sistema eléctrico venezolano (Red VEN)

La Figura 7 presenta las funciones aproximadas de densidad de probabilidades para cada una de las perturbaciones analizadas. De la figura resulta evidente que las variaciones en los pesos generan particiones diferentes, incluso en el caso de menor perturbación. Para variaciones del 5%, la probabilidad de que el índice v_i sea menor que 0,1 (esto es, la probabilidad de que las particiones que se generen sean iguales a la partición base), es de 0,7, aproximadamente.

A.3 Sistema IEEE 118 (Red IEEE)

La Figura 8 muestra las distribuciones de v_i para la red IEEE. Se aprecia un comportamiento similar al observado en las funciones de distribución de probabilidad del índice de comparación de particiones v_i en la red VEN. Para esta red, en el caso de variaciones del 5%, la probabilidad de generar particiones iguales a la partición base es aproximadamente 0,3.

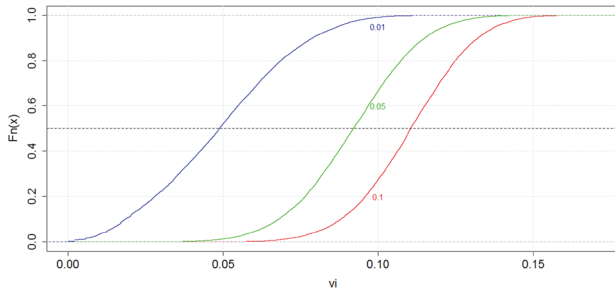


Figura 7. Función de distribución de probabilidad acumulada para el índice v_i de la red VEN, para diferentes ΔW_i

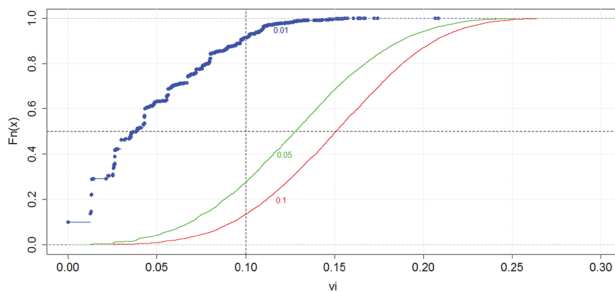


Figura 8. Función de distribución de probabilidad acumulada para el índice v_i de la red IEEE, para diferentes ΔW_i

A.4 Comparación de resultados

En general, para las redes eléctricas analizadas, al incrementarse el valor de la perturbación de los enlaces, el índice v_i aumenta, lo cual implica menor similitud en las particiones obtenidas respecto a la partición base.

A fin de comparar la red VEN y la IEEE, la Figura 9 presenta las distribuciones normalizadas de ambas redes, para la perturbación del 5 % (en el caso del 10 % el comportamiento es similar al presentado en la Figura 9).

Se puede concluir que los efectos de la incertidumbre de los pesos de los enlaces son mayores en la red IEEE que en la red VEN. La diferencia entre las dos redes eléctricas analizadas puede deberse a la topología de las redes ya que, en el caso de la red VEN, ésta presenta un mayor número de estructuras radiales.

Si se establece como límite máximo admisible $\text{Prob}(v_i \leq 0,1) \geq 0,6$, la red VEN cumpliría con este criterio ($\text{Prob}(v_i \leq 0,1) = 0,665$), pero no la red IEEE ($\text{Prob}(v_i \leq 0,1) = 0,286$). Bajo esta última situación, el analista pudiera revisar los valores límites asociados a la incertidumbre de los pesos de los enlaces, mediante un análisis de sensibilidad.

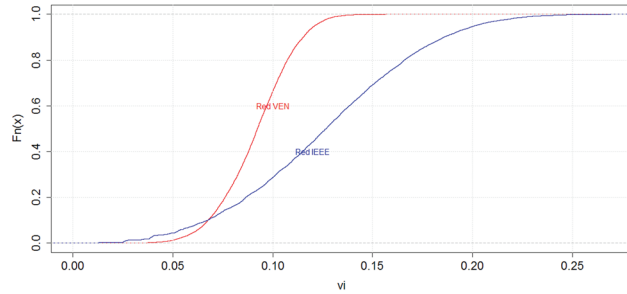


Figura 9. Comparación funciones de distribución, redes VEN vs IEEE ($\Delta W_i = 0,05$)

CONCLUSIONES

Este trabajo presenta una propuesta para analizar los efectos de las incertidumbres de los pesos de los enlaces en la detección de comunidades mediante el uso de índices de similitud entre particiones.

El procedimiento, basado en la simulación de Monte Carlo, permite obtener la función de distribución aproximada del índice de similitud seleccionado. El análisis de las distribuciones generadas para diferentes valores de incertidumbre, representa un valioso insumo para decidir y/o establecer la importancia de la incertidumbre de los pesos de los enlaces de una red, sobre todo, si las comunidades detectadas para los valores nominales van a ser utilizadas en estudios posteriores de vulnerabilidad.

La metodología propuesta, aplicable a cualquier red ponderada y cualquier método de detección de comunidades empleado, es ejemplificada mediante el uso del algoritmo Fastgreedy y el índice de comparación *Variation Information* (v_i).

La evaluación se realizó en tres redes: una red artificial con estructura comunitaria claramente definida; y dos sistemas eléctricos de potencia reales. Los resultados indican que la red artificial es robusta frente a las variaciones de los pesos de los enlaces. Sin embargo, para las redes eléctricas consideradas, la detección de comunidades es muy sensible al peso de los enlaces. En otras palabras, la probabilidad de obtener particiones iguales a la partición base es baja. Esta probabilidad disminuye a medida que las perturbaciones en los pesos de los enlaces aumentan.

La alta sensibilidad de las redes eléctricas a los pesos de los enlaces sugiere que probablemente las redes no posean estructuras comunitarias definidas claramente. De aquí que los estudios que se puedan derivar a partir de la detección de comunidades deban ser revisados en detalle para poder

definir valores límites apropiados en los pesos (ya sea que éstos representen, impedancia, confiabilidad, potencia transmitida, entre otros).

En futuros trabajos se pretende estudiar, mediante el uso de análisis de sensibilidad basado en descomposición de la varianza (Saltelli et al., 2000), cuál es el conjunto de enlaces que al ser perturbados originan la mayor diversidad en las particiones y proponer las medidas remediales adecuadas.

REFERENCIAS

- ALDECOA R. (2012). Detección de comunidades en redes complejas. Tesis de Maestría no publicada. Universidad Politécnica de Valencia, Valencia, España.
- BAGLER G. (2008). Analysis of the airport network of India as a complex weighted network. *Physica A: Statistical Mechanics and its Applications*. 387(12); pp 2972-2980.
- BOCCALETTI S., LATORA V., MORENO Y., CHAVEZ M., & HWANG D.U. (2006). Complex networks: Structure and dynamics. *Physics Reports*. 424; pp. 175-308.
- CLAUSET A., NEWMAN M. E., & MOORE C. (2004). Finding community structure in very large networks. *Physical Review E*. 70(6); pp 066111.
- DANON L., DIAZ-GUILERA A., DUCH J., & ARENAS, A. (2005). Comparing community structure identification. *Journal of Statistical Mechanics: Theory and Experiment*. 09; pp. 09008.
- FAN Y., LI M., ZHANG P., WU J., & DI Z. (2006). The role of weight on community structure of networks. arXiv preprint physics/0609218.
- FAN Y., LI M., ZHANG P., WU J., & DI Z. (2007). Accuracy and precision of methods for community identification in weighted networks. *Physica A: Statistical Mechanics and its Applications*. 377(1); pp. 363-372.
- FORTUNATO S. (2010). Community detection in graphs. *Physics Reports*. 486; pp. 75-174.
- GIRVAN M., & NEWMAN M. E. (2002). Community structure in social and biological networks. *Proceedings of the National Academy of Sciences*. 99; pp. 7821-7826.
- HAMAD I., RIKVOLD P., & POROSEVA S. (2011) Floridian high-voltage power-grid network partitioning and cluster optimization using simulated annealing. *Physics Procedia*. 15, pp. 2-6.
- HRIC D., DARST R. K., & FORTUNATO S. (2014). Community detection in networks: structural clusters versus ground truth. *Physical Rev. E*. 9; pp. 062805.
- HUBERT L., & ARABIE P. (1985). Comparing partitions. *Journal of classification*. 2(1); pp. 193-218.
- KARRER B., LEVINA E., & NEWMAN M. E. (2008). Robustness of community structure in networks. *Physical review E*. 77; pp. 046119.
- KLINDT W. (2000). Comparación de Técnicas para Evaluar el Efecto de Parámetros Imprecisos en un Modelo de Decisión. Trabajo de Grado de Maestría. Universidad Central de Venezuela, Caracas, Venezuela.
- MAROTTA L., MICCICHÈ S., FUJIWARA Y., IYETOMI H., AOYAMA H., GALLEGATI M., & MANTEGNA, R. N. (2014). Bank-firm credit network in Japan: An analysis of a bipartite network. e0123079.
- MEI S., ZHANG X., & CAO M. (2011). Power grid complexity. Springer Science & Business Media. 179-190.
- MEILÄ, M. (2003). Comparing clusterings by the variation of information. *Learning theory and kernel machines*. Springer Berlin Heidelberg. pp. 173-187.
- MORONTA J., & ROCCO C. (2016). Comparación de algoritmos de detección de comunidades en sistemas eléctricos de potencia. Artículo enviado a Revista de la Facultad de Ingeniería Universidad Central de Venezuela.
- NEWMAN M. E. (2004). Analysis of weighted networks. *Physical Review E*. 70(5); pp 056131.
- NEWMAN M. E. (2006). Modularity and community structure in networks. *Proceedings of the National Academy of Sciences*. 103; pp. 8577-8582.
- NEWMAN M. E., & GIRVAN M. (2004). Finding and evaluating community structure in networks. *Physical review E*. 69(2); pp. 026113.
- PESONEN J., & HYVÖVENN E. (1996). Interval Approach Challenges Monte Carlo Simulation. *Reliable Computing*. 2(2), pp. 155-160.
- POWER SYSTEMS TEST CASE ARCHIVE: University of Washington. En línea. Disponible en: <http://www.ee.washington.edu/research/pstca/>. Consultado: 02 de Junio de 2015.

- RAND W. M. (1971). Objective criteria for the evaluation of clustering methods. *Journal of the American Statistical association*. 66(336); pp. 846-850
- ROCCO C. (2008). Análisis del sistema eléctrico venezolano desde la perspectiva de la teoría de redes complejas. *Revista de la Facultad de Ingeniería Universidad Central de Venezuela*. 23(1); pp. 103-109.
- ROCCO C., & RAMIREZ-MARQUEZ J. (2011). Vulnerability metrics and analysis for communities in complex networks. *Reliability Engineering and System Safety*. 96; pp.1360–1366.
- SALTELLI, A., CHAN, K., SCOTT, E.. (2000). *Sensitivity Analysis*. John Wiley & Sons, Ltd The Atrium, Southern Gate, Chichester,, England.
- THE R PROJECT FOR STATISTICAL COMPUTING. (Consulta en noviembre, 2014). Disponible en: <http://www.r-project.org/>.
- ZACHARY, W. (1977). An information flow model for conflict and fission in small groups. *Journal of anthropological research*. pp. 452-473.