



Laboratorio
Interdisciplinario de
Estadística Social



Discussion Paper N°2017|04

Construcción de un Espacio de Conocimiento

Gabriel Muñoz y Rodrigo Vargas

Facultad de Matemáticas
Pontificia Universidad Católica de Chile
Av. Vicuña Mackenna 4860, Macul
lies.mat.uc.cl/



Laboratorio Interdisciplinario de Estadística Social

Para citar:

Muñoz, G. y Vargas, R. (2017). Construcción de un Espacio de Conocimiento. *LIES Discussion Paper* N°2017|04.

CONSTRUCCIÓN DE UN ESPACIO DE CONOCIMIENTO

G. Muñoz y R. Vargas.

INTRODUCCIÓN

Considere la tarea de evaluar el estado de conocimiento de un individuo con respecto a un conjunto específico de problemas o preguntas. Para concreción, se muestra una pequeña muestra de tales problemas a continuación, tomados de los planes y programas de matemáticas de primero y segundo medio. Estos serán los ítems que se utilizarán para el ejemplo

1. Multiplicar dos binomios.
2. Identificar y representar puntos en el plano cartesiano, manualmente o usando un procesador geométricos.¹
3. Resolver sistemas de ecuaciones lineales con dos incógnitas, gráfica y algebraicamente.²
4. Fracciones equivalentes.³
5. Analizar representaciones de la función lineal y de la función afín.⁴

Los examinadores podrían hacer los cinco problemas y examinar las respuestas. Sin embargo, tal método sería ineficaz e impráctico si tienen 10.000 estudiantes para examinar y la prueba tiene 80 problemas. La eficiencia se puede mejorar aprovechando una estructura implícita que relaciona estos problemas. Por ejemplo, podemos deducir con seguridad que cualquier estudiante capaz de resolver el Problema 3 puede también resolver los problemas 1, 4 y 5 (verificaremos esto experimentalmente más adelante). Por lo tanto, hacer estos tres problemas sería una pérdida de tiempo. Del mismo modo, un estudiante que no responde correctamente el Problema 4 casi seguramente no responderá correctamente los Problemas 1 o 3. Antes de proceder, eliminemos algunos posibles malentendidos. No estamos asumiendo que la estructura es necesariamente simple (por ejemplo, hay un orden de dificultad sobre los problemas, tales como $2 < 4 < 1 < 5 < 3$), ni estamos sugiriendo que la estructura simplemente depende de las relaciones lógicas entre los problemas.

En general, observamos que, a partir de las respuestas (correctas o incorrectas) a algunos problemas en un campo dado, algunas inferencias probables pueden hacerse con respecto a las respuestas a otros problemas en ese campo; estas inferencias podrían utilizarse en el diseño de procedimientos eficaces de evaluación de conocimientos. Obviamente, tenemos procedimientos informáticos en mente.

El concepto central en nuestro estudio es el de un estado del conocimiento, por lo que nos referimos a un conjunto que contiene todos los problemas que algún individuo es capaz de resolver. La colección de todos estos estados se denomina estructura de conocimiento. No se adjunta interpretación cognitiva

¹Primero medio, Geometría. Aprendizaje Esperado 01

²Segundo medio, Álgebra. Aprendizaje Esperado 06

³Primero medio, Números. Aprendizaje Esperado 07

⁴Primero medio, Álgebra. Aprendizaje Esperado 04

a estos conceptos. Sin embargo, se puede demostrar que son consistentes con algunas características estándar de la teoría psicométrica, como habilidades o destrezas.

CONCEPTOS BÁSICOS

Inicialmente, las preguntas que nos hicimos fueron las siguientes: ¿Cómo podemos utilizar las respuestas anteriores dadas por un sujeto a algunos problemas para identificar los estados de conocimiento posibles restantes? ¿Cómo podemos usar esa información para elegir el próximo problema para dar al sujeto? Más fundamental, ¿cómo podemos estructurar el conjunto de problemas para permitir tales inferencias, y cómo se puede definir el concepto de un estado del conocimiento?

Denotamos por Q el conjunto completo de problemas bajo consideración. En nuestro ejemplo, este conjunto contiene Problemas 1, 2, 3, 4 y 5. En la práctica, el conjunto Q puede ser bastante grande y contener docenas, tal vez cientos de problemas. Como se indica, suponemos que las respuestas se codifican sólo como correctas o incorrectas. La primera formalidad que viene a la mente se basa en la idea de que, desde la observación de que un estudiante es capaz de resolver un problema dado, a veces se puede conjeturar que este estudiante también puede resolver otros problemas. Esto sugiere la introducción de una relación binaria $\lesssim$, con la siguiente interpretación:

$$q \lesssim t \iff \begin{array}{l} \text{de una respuesta correcta al problema } t; \\ \text{podemos suponer una respuesta correcta al problema } q. \end{array}$$

Acortar esto como q se puede deducir de t , o bien se dice que q precede a t . La relación $\lesssim$ se llamará la relación de precedencia. Esta relación es obviamente reflexiva, y es razonable suponer que también es transitiva. En otras palabras, es un cuasi-orden en el conjunto Q .

Una relación de precedencia $\lesssim$ para nuestros cinco problemas se muestra en la Figura 1. Las convenciones de la figura (conocido como un diagrama de Hasse en la teoría de los grafos) son los siguientes. Cuando se puede llegar a algún problema q desde algún otro problema t por líneas descendentes, esto significa q se puede deducir de t ; Es decir, $q \lesssim t$.

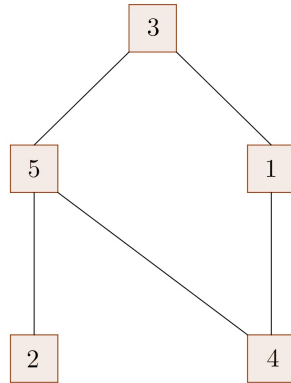


Figura 1: El diagrama de Hasse de la relación de precedencia de los cinco problemas.

Por ejemplo, a partir de una respuesta correcta al problema 2, podemos suponer una respuesta correcta a los problemas 5 y 3 (por lo tanto, $2 \lesssim 5$, $5 \lesssim 3$ y $2 \lesssim 3$). De esta representación se puede deducir una lista completa de los posibles estados de conocimiento. El estudiante puede no saber nada, o puede saber sólo el problema 2, o simplemente el problema 4, o quizás los problemas 1 y 4, o los problemas 1,2 y 4, etcetera. Hay 9 estados de conocimiento de un posible $2^5 = 32$:

$$\emptyset, \{2\}, \{4\}, \{1, 4\}, \{2, 4\}, \{1, 2, 4\}, \{2, 4, 5\}, \{1, 2, 4, 5\} \quad \text{y} \quad \{1, 2, 3, 4, 5\}.$$

Formalmente, la definición del estado de conocimiento correspondiente a la relación de precedencia $\lesssim$ es

$$K \subseteq Q \text{ es un estado} \iff \text{para todo } q, t \in Q, q \lesssim t, \text{ y } t \in K \Rightarrow q \in K.$$

En un lenguaje sencillo: Un conjunto K de problemas es llamado un estado si, cuando K contiene algún problema t , también contiene todos los problemas que se pueden deducir de t . Hay un rasgo accidental de nuestro ejemplo: No hay caso en el que dos de los problemas pueden ser conjeturados el uno del otro. Formalmente, esto significa que la relación de precedencia $\lesssim$ satisface la condición de antisimetría $q \lesssim t$ y $t \lesssim q \Rightarrow q = t$. Un cuasi-orden que también es antisimétrico se denomina un orden parcial.

Sin embargo, apelando a la simplicidad de esta formalización, tiene una consecuencia indeseable: Cualquier problema t tiene un conjunto único de prerrequisitos o antecedentes, a saber, el conjunto de todos los problemas que se pueden conjeturar desde t . En el diagrama de Hasse, estos problemas pueden ser alcanzados desde (por líneas descendentes, por ejemplo, el conjunto de prerrequisitos del Problema 5 es el conjunto $\{2, 4\}$). Esta condición puede ser considerada aceptable en nuestro ejemplo, pero es ciertamente absurda en general. Los contraejemplos son fáciles de fabricar, incluso en un campo muy estructurado como las matemáticas. Supongamos que uno descubre que un estudiante es capaz de resolver un sistema de ecuaciones lineales, por ejemplo, tres ecuaciones en tres incógnitas. Se puede deducir de este hecho que el estudiante conoce el concepto de determinante o que el estudiante es capaz de invertir una matriz, o que el estudiante sabe cómo manipular ecuaciones (sumarlas o restarlas) como en el método de Gauss. Además, cualquier combinación de estas tres hipótesis es posible. Cualquiera que sea el caso no se determina.

Por lo tanto, el modelo es demasiado fuerte. Sin embargo, qué suposición o suposiciones deben ser dejadas de lado no está clara. En esta coyuntura, un cambio de punto de vista será rentable. Lo que se conservará es la idea fundamental de que el estado de conocimiento de algún individuo es el conjunto de todos los problemas que este individuo es capaz de resolver (omitiendo, por el momento, complicando factores como respuestas incorrectas debido a errores descuidados y respuestas correctas resultantes de suposiciones con suerte). La estructura del conocimiento es entonces la colección de todos los subconjuntos de problemas que representan estados. Así, en nuestro ejemplo, denota por $\mathcal{K}$ la estructura de conocimiento, obtenemos

$$\mathcal{K} = \{\emptyset, \{2\}, \{4\}, \{1, 4\}, \{2, 4\}, \{1, 2, 4\}, \{2, 4, 5\}, \{1, 2, 4, 5\}, \{1, 2, 3, 4, 5\}\}. \quad (1)$$

Claramente, no todos los subconjuntos de problemas son adecuados para representar estados. Para determinar los tipos de restricciones que pueden surgir

para una estructura de conocimiento, uno puede preguntarse: ¿Hay algo especial acerca de la estructura del conocimiento derivada de una relación de precedencia que resulta ser un cuasi-orden, como el representado en la Figura 1? Observe que la colección de estados especificados por la ecuación (1) satisface la propiedad siguiente:

Si K y K' son estados, entonces $K \cup K'$ es también un estado.

Por ejemplo, tenemos los dos estados $\{2, 4\}$ y $\{1, 4\}$, pero también tenemos el estado $\{1, 2, 4\} = \{2, 4\} \cup \{1, 4\}$. Una familia de estados que satisfacen esta propiedad será llamada unión cerrada, o $\cup$ -cerrada. Otra propiedad, en el mismo sentido pero que implica intersecciones, requiere que

Si K y K' son estados, entonces $K \cap K'$ es también un estado.

Si esta propiedad es válida para una familia de estados, decimos que es intersección cerrada, o $\cap$ -cerrada. De la ecuación (1), es fácil verificar que la familia de estados $\mathcal{K}$ de nuestro ejemplo es $\cup$ -cerrada y $\cap$ -cerrada. Este ejemplo ilustra una situación general, que puede describirse en términos del siguiente resultado clásico.

Teorema 1 (Birhhoff, 1937) *Para cualquier conjunto Q , la fórmula*

$$q \lesssim t \Leftrightarrow (t \in K \Rightarrow q \in K, \text{ para todo } K \in \mathcal{K}) \quad (2)$$

Establece una correspondencia uno-a-uno entre el conjunto de todos los cuasi-ordenes $\lesssim$ en Q , y la colección de todas las familias $\mathcal{K}$ de los subconjuntos de Q que son $\cup$ -cerrado y $\cap$ -cerrado.

En otras palabras: Para cualquier relación de precedencia $\lesssim$ sobre el conjunto Q , que es reflexiva y transitiva, hay exactamente una familia de estados $\mathcal{K}$ que es $\cup$ -cerrada y $\cap$ -cerrada, y viceversa. Además, las especificaciones de $\mathcal{K}$ por $\lesssim$ y de $\lesssim$ por $\mathcal{K}$ están dadas por la Fórmula (2).

Este resultado es útil en la medida en que introduce una perspectiva muy diferente en una colección de estados de conocimiento. Anteriormente, la condición definitoria era que una colección de estados tenía que ser consistente con (o especificar en términos de) una relación de precedencia, que se supone era transitiva y reflexiva. Ahora, sin embargo, se puede prescindir de la relación de precedencia.

Nuestra discusión se resume en la siguiente definición fundamental

Definición 1 *Una estructura de conocimiento es un par $(Q, \mathcal{K})$ en el cual Q es un conjunto de problemas o preguntas y $\mathcal{K}$ es una colección de subconjuntos de Q , llamados los estados de conocimiento.*

Ocasionalmente, cuando no puede surgir ambigüedad en relación con el conjunto de problemas, nos referimos a $\mathcal{K}$, en lugar de a $(Q, \mathcal{K})$, como la estructura del conocimiento. El conjunto de todos los estados que contienen un problema particular q será denotado por $\mathcal{K}_q$. El conjunto de todos los problemas pertenecientes a los mismos estados que q se denota por $[q]$ y se llama una noción. Por lo tanto,

$$[q] = \{r \mid \mathcal{K}_r = \mathcal{K}_q\}.$$

Se dice que una estructura de conocimiento es discrimina cuando cada noción contiene exactamente un problema, es decir, cuando $[q] = \{q\}$ para todos los problemas q . (Es el caso de la estructura de conocimiento $\mathcal{K}$.)

Definición 2 Una estructura de conocimiento $(Q, \mathcal{K})$ se denomina espacio de conocimiento cuando los dos axiomas siguientes son válidos:

[K1] El conjunto Q y el conjunto vacío $\emptyset$ son estados.

[K2] Cualquier unión de estados es un estado (esto es, $\mathcal{K}$ es $\cup$ -cerrado).

Suponga que, adicionalmente, la estructura de conocimiento satisface

[K3] Cualquier intersección de estados es un estado (esto es, $\mathcal{K}$ es $\cap$ -cerrado).

Entonces, $\mathcal{K}$ es llamado un espacio cuasi-ordinal. Un espacio cuasi-ordinal y discriminativo es un espacio parcialmente ordinal. (Es fácil demostrar que cuando una estructura de conocimiento $\mathcal{K}$ es definida a partir de un cuasi-orden $\lesssim$, entonces $\mathcal{K}$ es discriminativa si y sólo si $\lesssim$ es un orden parcial, esto es, $\lesssim$ es antisimétrica).

La estructura de conocimiento $\mathcal{K}$ es un espacio parcialmente ordinal cuando la ecuación (1) define una estructura discriminativa, $\cup$ -cerrado y $\cap$ -cerrado.

Una representación gráfica del espacio de conocimiento definido por la Ecuación (1) se da en la Figura 2.

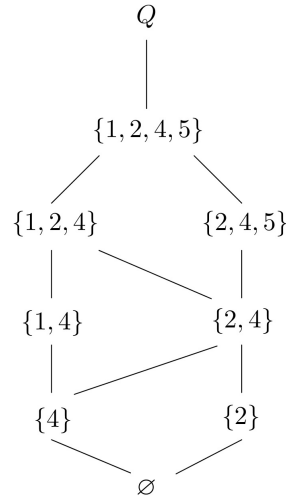


Figura 2: Gráfico del espacio de conocimiento de la ecuación (1)

¿Cómo construir un espacio de conocimiento en la práctica? A nuestro juicio, un primer esbozo del espacio debe obtenerse consultando sistemáticamente a maestros con experiencia. Sin embargo, no podemos simplemente pedir a estos expertos que nos den una lista de los posibles estados de conocimiento. Por una razón, el concepto de un estado del conocimiento es algo abstracto y puede ser difícil de expresar exactamente. Por otra parte, tal lista ciertamente contendrá, en cualquier aplicación realista, miles de estados. Por lo tanto, ¿cuáles son las preguntas correctas para pedir a los expertos? Esbozamos un procedimiento práctico, que ilustramos mediante una aplicación que involucra a nuestros cinco problemas y el espacio de conocimiento representado en la Figura 2.

CONSTRUYENDO EL ESPACIO DE CONOCIMIENTO

Incluso para conjuntos de tamaño moderado (por ejemplo, Q contiene 30 problemas), la construcción práctica del espacio de conocimiento es una empresa considerable. Esta construcción debe basarse en parte en la experimentación, lo que significa recolectar las frecuencias de aparición de todos los posibles patrones de respuestas en una gran muestra de sujetos y, a través de algún análisis apropiado, inferir el espacio de conocimiento que mejor pueda dar cuenta de los datos. Hay mucho que ganar al obtener primero un bosquejo del espacio del conocimiento basado en las opiniones de los profesores expertos. Sin embargo, como ya se ha argumentado, preguntar a un profesor, no importa cuán experimentado, si determinados conjuntos de problemas son posibles estados de conocimiento no parece factible. Debe adoptarse un enfoque indirecto.

Hay, sin embargo, un tipo de pregunta que un experto comprenderá plenamente y podría también ser capaz de responder confiablemente:

[Q1] Supongamos que un estudiante bajo examen acaba de proporcionar una respuesta incorrecta al problema q_1 . ¿Es prácticamente cierto que este estudiante también fallará el problema q_2 ? Suponemos que se excluyen los errores por descuidos y las conjeturas afortunadas.

Para propósitos expositivos, simplificamos esto a la pregunta: ¿Fallando q_1 implica fallar q_2 ?

El maestro puede sacar de su experiencia educativa con una población de estudiantes para decidir si responder “sí” o “no”. Una respuesta positiva significaría que algunos conjuntos de problemas pueden ser eliminados como estados posibles, es decir, todos aquellos conjuntos que contienen el problema q_2 pero no el problema q_1 . De hecho, si hay algunos estados en el espacio de conocimiento que contiene q_2 y no q_1 , entonces podría haber un estudiante que falla q_1 , pero no q_2 , y el experto debe responder “no” a [Q1], una respuesta positiva a [Q1] se codificará como $q_1 P q_2$, y una respuesta negativa se codificará como $q_1 N q_2$. Tenga en cuenta que no todas estas preguntas deben hacerse. Por ejemplo, si hemos observado $q_1 P q_2$ y $q_2 P q_3$, podemos inferir $q_1 P q_3$ (si fallar q_1 implica fallar q_2 y fallar q_2 implica fallar q_3 , entonces, seguramente, fallar q_1 , implica fallar q_3). Luego, no es necesario preguntar [Q1] para el par (q_1, q_3) . Lo que significa que P , considerada como una relación binaria, es transitiva sobre el conjunto de problemas Q . Esta propiedad de transitividad puede ser usada para facilitar considerablemente la tarea del experto. Lamentablemente, si las únicas preguntas que se hicieran fueran del tipo [Q1], sólo podríamos construir el espacio del conocimiento en algunos casos muy especiales, a saber, casos en los que el espacio es $\cup$ -cerrado y $\cap$ -cerrado (esto se deduce del teorema de Birkhoff). Es evidente que puede faltar alguna información.

Esta discusión sugiere un tipo más general de pregunta para adquirir información del experto:

[Q2] Supongamos que un estudiante bajo examen ha proporcionado respuestas equivocadas a todos los problemas en un subconjunto A de Q . ¿Es prácticamente cierto que este estudiante también fallará el problema q en Q ?
Asumimos que los errores por descuidos y suposiciones con suerte están excluidos.

Extendiendo nuestra notación, codificaremos respuestas positivas y negativas a preguntas tales como APq y ANq , respectivamente. Una respuesta positiva a la pregunta: ¿El fracaso de 1 y 3 implica fallar 2? Será denotado por $\{1, 3\}P2$. Se puede establecer matemáticamente que, si todas las preguntas de tipo [Q2] se les pide, entonces el espacio de conocimiento puede ser recuperado exactamente.

Nuestro primer paso consiste en ordenar las preguntas que deben hacerse a los expertos; es decir, tenemos que ordenar los pares (A, q) que corresponden a las preguntas: ¿falla todos los problemas en A implica fallar el problema q ?, es razonable comenzar con las preguntas más simples. El orden de preguntas por el tamaño del conjunto A en el par (A, q) . Luego, decimos que (A, q) está antes que (A', q') cuando $|A| < |A'|$. (Aquí, y en lo que sigue, denotamos por $|S|$ el número de elementos en un conjunto S .) Esto significa que todas las preguntas de tipo [Q1] serán preguntadas en la primera etapa del procedimiento.

Observe que si $q \in A$ entonces APq , el conteo de estas respuestas positivas a priori es fácil de hacer. Si el número de problemas en Q es n (es decir, si $|Q| = n$) entonces el número de este tipo de preguntas es $2^{n-1} \times n$. En nuestro caso, con 5 problemas se tiene 80 preguntas positivas, que no es necesario realizar. El número de preguntas de la forma ARq con $|A| = 1$ es $(n-1) \times n$ y el número de preguntas de la forma ARq con $|A| = k > 1$ es $\binom{n}{k} \times (n-k)$. En total, el número de preguntas necesarias para determinar el espacio de conocimiento es $(2^{n-1} - 1) \times n$, sin embargo usando las inferencias debidas a la transitividad de la relación de precedencia y al teorema de Birkhoff es posible realizar un número menor de preguntas. En nuestro ejemplo, se tuvieron que realizar 36 preguntas para determinar el espacio de conocimiento de un total de 75 preguntas.

En la siguiente tabla se tiene que la primera columna indica el número de paso; la segunda columna enumera la respuesta observada en el número de paso actual; la tercera columna contiene las inferencias negativas extraídas de la respuesta observada (incluida la respuesta en sí, si es negativa); la cuarta columna contiene las inferencias positivas extraídas de la respuesta observada (incluida la respuesta en sí); la quinta columna muestra qué estados se suprimen del espacio de conocimiento actual si se obtiene una respuesta positiva; la sexta columna contiene el número actual de respuestas negativas e inferencias; y la séptima columna contiene el número actual de respuestas positivas e inferencias. Los guiones indican que no se agrega ni se elimina nada.

Tabla 1
Pasos del procedimiento para interrogar al experto

Paso	Respuesta	Pares agregados a N_t	Pares agregados a P_t	Estados de conomiento eliminados de $\mathcal{K}$	$ N_t $	$ P_t $
1	1N2	1N2	-	-	1	80
2	1P3	3N2, {1,3}N2	{1,2}P3, {1,4}P3, {1,5}P3, {1,2,4}P3, {1,2,5}P3, {1,4,5}P3, {1,2,4,5}P3	{3}, {2,3}, {3,4}, {3,5}, {2,3,4}, {2,3,5}, {3,4,5}, {2,3,4,5}	3	88
3	1N4	1N4	-	-	4	88
4	1N5	1N5	-	-	5	88
5	2N1	2N1	-	-	6	88
6	2P3	3N1, {2,3}N1	2P3, {2,4}P3, {2,5}P3, {2,4,5}P3,	{1,3}, {1,3,4}, {1,3,5}, {1,3,4,5}	8	92
7	2N4	2N4	-	-	9	92
8	2P5	5N4, 5N1, {2,5}N4, {2,5}N1	2P5, {1,2}P5, {2,3}P5, {2,4}P5, {1,2,3}P5, {1,2,4}P5, {2,3,4}P5, {1,2,3,4}P5	{5}, {1,5}, {4,5}, {1,4,5}	13	100
9	3N4	3N4	-	-	14	100
10	3N5	3N5	-	-	15	100
11	4P1	-	4P1, {2,4}P1, {3,4}P1, {4,5}P1, {2,3,4}P1, {3,4,5}P1, {2,4,5}P1, {2,3,4,5}P1	{1}, {1,2}, {1,2,3}, {1,2,5}, {1,2,3,5}	15	108
12	4N2	4N2	-	-	16	108
13	4P3	{3,4}N2	4P3, {4,5}P3	-	17	110
14	4P5	5N2, {4,5}N2	4P5, {1,4}P5, {3,4}P5, {1,3,4}P5	{2,5}	19	114
15	5P3	-	5P3	{1,2,3,4}	19	115
16	{1,2}N4	{1,2}N4	-	-	20	115
17	{1,3}N4	{1,3}N4	-	-	21	115
18	{1,3}N5	{1,3}N5	-	-	22	115
19	{1,4}N2	{1,4}N2	-	-	23	115
20	{2,3}N4	{2,3}N4	-	-	24	115
21	{3,5}N1	{3,5}N1	-	-	25	115
22	{3,5}N2	{3,5}N2	-	-	26	115
23	{3,5}N4	{3,5}N4	-	-	27	115
24	{1,5}N2	{1,5}N2	-	-	28	115
25	{1,5}N4	{1,5}N4	-	-	29	115
26	{1,2,3}N4	{1,2,3}N4	-	-	30	115
27	{1,2,5}N4	{1,2,5}N4	-	-	31	115
28	{2,3,5}N1	{2,3,5}N1	-	-	32	115
29	{2,3,5}N4	{2,3,5}N4	-	-	33	115
30	{3,4,5}N2	{3,4,5}N2	-	-	34	115
31	{1,3,4}N2	{1,3,4}N2	-	-	35	115
32	{1,4,5}N2	{1,4,5}N2	-	-	36	115
33	{1,3,5}N2	{1,3,5}N2	-	-	37	115
34	{1,3,5}P4	{1,3,5}P4	-	-	38	115
35	{1,3,4,5}N2	{1,3,4,5}N2	-	-	39	115
36	{1,2,3,5}N4	{1,2,3,5}N4	-	-	40	115